By How Much Did Railroads Conquer the West?

Ross Knippenberg*
University of Colorado at Boulder

September 5, 2013

Abstract

I study the effect of railroads on integrating the economy of the United States in the mid- to late-nineteenth century. I have assembled a unique data set using hand-collected price data on 13 commodities for 283 locations in the United States over the years 1851-1892. A railroad connection between two places should reduce transportation costs and thus price differences. I use the variation in prices, geography and railroad networks to answer a variety of questions about the role of the railroads in market integration. Recognizing that railroads were built for profit and not at random, a railroad connection may be correlated with price gaps. To correct for this, I use an instrumental variables approach where population, terrain ruggedness and a function of elevation instrument for whether a city is connected to the railroad. Using this estimation method, I find that a railroad connection tended to reduce the price gap between two cities by an average of 70 percent.

Keywords: History of Railroads, Railways, Trains, Transportation, Infrastructure, Transcontinental Railroad, Prices, Price Dispersion, Old West

JEL No. N71, N91, F14

Acknowledgements

I would like to thank my advisor, Murat Iyigun for guidance throughout this project, members of my dissertation committee: Lee Alston, Carol Shiue and Carlos Martins-Filho, as well as David Atkin, Dave Donaldson, Robert Margo, Jeremy Atack, Matt Jaremski, Ann Carlos, Marta Felis-Rota, Melinda Miller, Paul Rhode, John Wallis, Federico Giovanni, Rick Steckel, Yukiko Abe, Farley Grubb, Larry Neal, Catherine Massey, Lee Chambers, Pete Veru, participants of the 7^{th} World Congress of Cliometrics and the 2012 Rocky Mountain Interdisciplinary History Conference for useful comments and discussion. All mistakes are my own.

*Department of Economics, University of Colorado at Boulder. Email: `Ross.Knippenberg@colorado.edu`. Web: `https://sites.google.com/site/rossknippenberg/`.

1 Introduction and Literature

Newspaper editor C.M. Chase, while in Leadville, Colorado, observed that: "Prices in Leadville are not what they were before the advent of the railroad, but they are sufficient to stagger an Eastern man, even now. What would you think, reader, to have to pay $46 a ton for hay, and not a prime article at that. How would you like to eat Irish potatoes at $2.50 a bushel, or give $1.80 a bushel for corn for your horse? Take it or leave it, there are plenty of calls for it on those terms.", (Chase, 1881, pg.179). This excerpt begs the question: just how important was the railroad, on average, in reducing price dispersion between areas of the United States? It is important to study prices near the beginning of the railroad: the mid to late 1800s was a time of great change with the spreading of the railroad. To put a number on the magnitude of the railroad over this period, consider that Fishlow (1966) for 1849-1890 estimated that railroad passenger costs were cut a quarter, freight rates were cut by three-quarters, railroad output grew by 100-fold, capital stock grew by 30-fold, employment grew from 18,000 to 750,000, and total factor productivity grew 2-to-3 fold (pgs 585, 606, 613, 626).

Table 1: Miles of Railroad

Year	Miles of Rails	Year	Miles of Rails
1851	10982	1872	66171
1852	12908	1873	70268
1853	15360	1874	72385
1854	16720	1875	74096
1855	18374	1876	76808
1856	22016	1877	79082
1857	24503	1878	81747
1858	26968	1879	86556
1859	28789	1880	93262
1860	30626	1881	103108
1861	31286	1882	114677
1862	32120	1883	121422
1863	33170	1884	125345
1864	33908	1885	128368
1865	35085	1886	136386
1866	36801	1887	149262
1867	39250	1888	156164
1868	42229	1889	161326
1869	46844	1890	166703
1870	52922	1891	170779
1871	60293	1892	175220

Source: 1900 Statistical Atlas, table number 100, page 374. Originally from Poor's Railroad Manual.

In this paper I examine the effect that railroads had on market integration in the U.S. with special emphasis on cities in the West. The regression results show a large decrease in bilateral price gaps when railroads are introduced to a city, an average of 70 percent. This is achieved by railroads integrating markets, decreasing the effects of local supply and demand shocks, decreasing local monopoly power, and enabling consumers to consume a wider range of goods. My contribution is that I am the first to examine the effect of railroads in the United States in terms of price dispersion.

Railroads are a major mover of goods today, and were even more important when first constructed in the 19th century. Railroad companies were the first large corporations and were a direct application of the mechanical technological revolution. At times the railroads created urban population centers, especially along the transcontinental routes. Learning the effects of the railroad are important because of their effect on infrastructure. Additionally, early railroads were a great experiment in corporations, monopolies, and adoption of technology. In my sample period, from 1851 to 1892, railway lines grew from 10,982 miles to 175,220 miles, (see Table 1) according to the 1900 Census Statistical Abstract (page 374). At an average of about \$60,000 per mile (page 377), railroads were easily the most expensive endeavor in the United States at the time.

In the Economic History literature, Fogel (1964) is the logical starting place whenever talking about the impact of railroads on the economy. His headline estimate was that a lack of railroads would reduce Gross Domestic Product by a mere 2.8 percent. Subsequent scholars have continuously revised and refuted this number, with the most recent being Donaldson and Hornbeck (2011), estimating the effect at 6.3 to 5.8 percent. These papers and many other like them take a market access approach to estimating the effect of railroads on the value of agricultural land.

Papers taking a price approach to estimating the effect of railroads include Donaldson (2012), Keller and Shiue (2008), and Jacks (2006). Donaldson (2012) studies the effect of the railroad in India and finds that it decreased transportation costs by 73 percent, among his other results. Keller and Shiue (2008) study the relative effects of railroads and the Zollverein customs union in Europe and find that trains reduced price gaps by about 14 percent, while reduced trade barriers decreased price gaps by 6 to 7 percent. Methodologically, my paper is most similar to Keller and Shiue (2008), particularly in the way they use population as an instrumental variable for whether a city gets a railroad connection. I add evidence to these two diverging estimates, finding that the choice of commodity has a huge impact on the estimate of the change in price dispersion: Donaldson uses salt from a particular place in India, while Keller and Shiue use wheat prices, and wheat is grown in every region of their study. Jacks (2006) takes a sample of cities in the U.S. and Europe finds little effect from railroad connections. He finds that railroads reduce trade costs by only 6.8 percent, even though they increase adjustment speed by 25.9 percent. He concludes that changes in monetary regimes and commercial policy appear to have a much greater effect on prices than do changes in transportation technology.

Additional recent papers examining railroads in the U.S. but not taking a price approach is Atack, Bateman, Haines, and Margo (2009). The authors explore the effects of railroads on urbanization, and, using government records of proposed railroad routes from 1824 to 1838 (without regard to whether or not they were actually built) as an instrument for a railroad connection [1] , they find that railroads doubled the rate of urbanization in the Midwest. Lastly, Atack, Jaremski and Rousseau (2013) examine the effect of railroads on the formation of banks in the Midwest before the Civil War. They find that railroads indeed increased the presence of banks in the Midwest.

2 Theoretical Framework

In a Hotelling Line Model, two retailers sell a homogeneous good at two different locations. A consumer is located somewhere between the two retailers. On a continuous number line, the consumer has probability zero of being located exactly in the middle between the two firms, so transportation costs are not equal, allowing each firm to set a different price. This creates a quasi-monopoly for each firm, where the firms maximize profits by setting their prices so that the consumer is indifferent between either firm. Thus the distance between markets is what creates differences in prices, or price gaps, and over time this leads to predictable patterns of price dispersion. In the Hotelling Spokes Model, as in Chen and Riordan (2007), the same idea holds for more than two firms. Again, the location of the consumer, or distance between markets, is what creates the price dispersion. These two models show that as the distance between markets increases, then price dispersion should also increase. Applying this idea to the railroads, the railroads caused a large decline in transportation costs, which should lead to a decline in price dispersion.

3 Data

I have constructed an original dataset for this paper using hand-collected data from multiple sources. The dependent variable is the log of the price gap between two merchants for each commodity. I have data from 283 merchants, some of which are located in the same city, especially large cities like New York, Boston, or St. Louis. Observations are by merchant or merchant-pairs. The thirteen commodities are superfine wheat flour, salted (or mess) pork, bacon, fresh beef, rice, soap, flannel cloth, satinets, brown sheeting, salt, black tea, green coffee from Rio de Janiero, and white navy beans. The black tea and green coffee are strictly imports. Soap, flannel, satinets, and brown sheeting are manufactured goods. One could group salt in with them as well, but I do not. The rest are all processed and homogenized agricultural commodities, produced in many parts of the country, except rice which is primarily produced in South Carolina or else imported from India.

The advantage of using prices for multiple commodities is that technological developments can

[1] Similar to the instrument used in Donaldson (2012) and Bannerjee, Duflo, and Qian (2006)

heterogeneously affect any particular product. For example, Olmstead and Rhode (2002) detail the evolution of wheat varieties on geographical productivity levels, namely that new wheat varieties drastically increased production and opened up millions of acres to wheat production which would otherwise have been left fallow. Secondly, refrigerated cars were invented during the sample period, and substantially decreased the price of selling fresh beef. As for the data sources, price levels for San Francisco come from Berry (1983). The Aldrich Report (Aldrich, 1892) reports price series of varying composition and length for 11 army posts, 13 naval bases, and 140 Indian agencies and schools. However, the vast majority of 1851-1880 prices comes from the Weeks Report (Weeks, 1883), which covers about 95 retail establishments and is the earliest known national survey of retail prices in the U.S. I deflate all prices using the deflators calculated in Warren and Pearson (1939) [2]. Table 2 describes some summary statistics of the data an Figure 1 shows the geographic distribution of the data.

The sample data is understandably full of holes. First of all no single merchant or price data source carries all of the goods in the sample. Secondly, some merchants only come into existence or have data available until after the railroad arrives to its city. In all I have 1,625 time series of prices. 10 percent of these series are for merchants who never connect to the railroad. It's worth mentioning that most of these are Indian Agencies in the West, some of which have poor data coverage. 68 percent of the series begin at or after the railroad connects. No city in my sample unconnects from the railroad during this time period. Lastly, 22 percent of the series are "switchers," meaning that I have price data both before and after connecting to the railroad.

Further details on the dataset can be found in the Data Appendix at the end of this paper.

4 Empirical Implementation

I consider three approaches to estimating the effect of railroads on prices, but only with the last one do I make a causal interpretation. The first two methods do not yield conclusive results but nontheless show interesting correlations. The first empirical exercise is to run a panel regression comparing the price levels of connected versus unconnected merchants. While this is not an estimate of the effect of the railroads, it is interesting enough to discuss briefly. Secondly, I roughly follow the border-effect methodology of Engel and Rogers (1996) by comparing the difference in price correlations before and after two merchants are connected to the railroad. The third method closely follows Keller and Shiue (2008) by taking a differences-in-differences approach: comparing price gaps before versus after connecting to the railroad. Connecting to the railroads is an endogenous variable, so I instrument for it with a combination of variables taken from previous literature. I

[2] Available from the NBER Macrohistory database. Earlier versions of this paper used the deflators calcualted in Coelho and Shepherd (1974), and Hoover (1958, 1960), but these deflators were solely based off of the Weeks data and thus only covered the period 1851 to 1880.

Table 2: Price Levels Summary Statistics (13,834 obs)

Variable	Units	Mean	Min	Max
Price	¢/pound	19.2	0.184	2212
Year	years	1871	1851	1892
Connected	dummy	88.4%	0	1
West	dummy	22.8%	0	1
Coast	dummy	23.4%	0	1
River	dummy	53.5%	0	1
Manufactured	dummy	26.9%	0	1
1870 Population	1000 people	113	0	942
1880 Population	1000 people	145	.033	1206
1890 Population	1000 people	200	.016	1647
Ruggedness	avg Δ meters	12.60	0.375	105.25
Elevation	meters	191.3	1	2714
Rainfall	cm/year	1024.8	70	1557

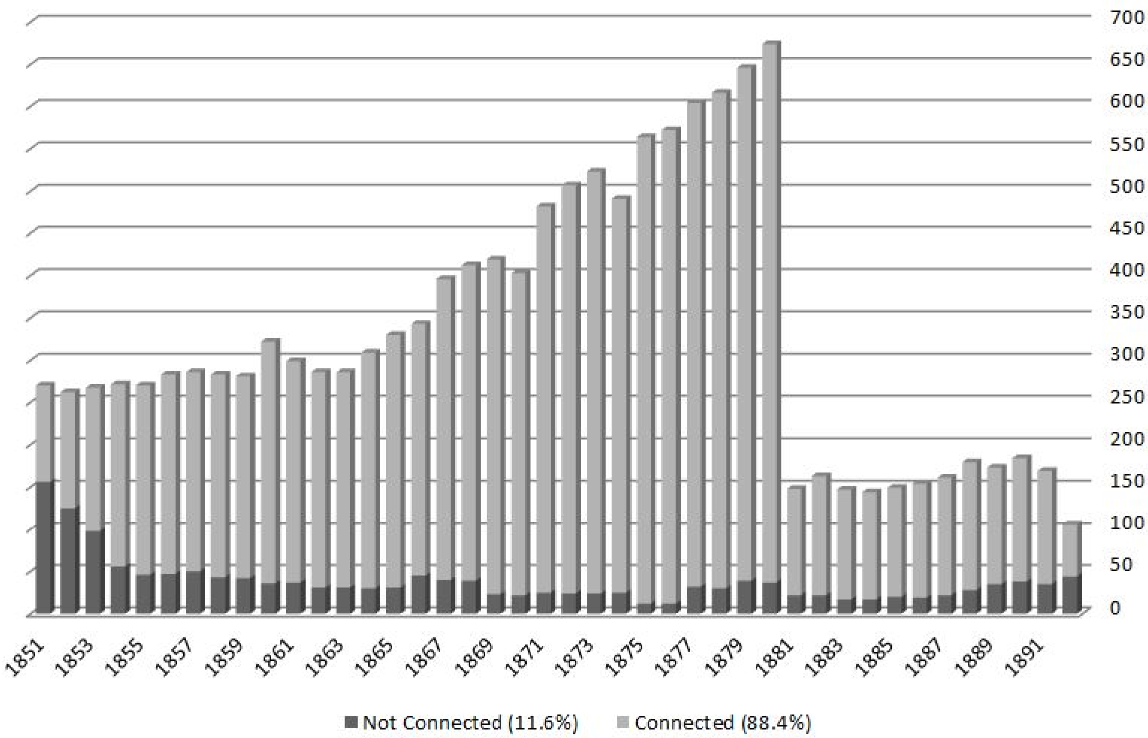
Figure 1: Geographic Distribution of the Price Data

Size of dot indicates the number of observations: 1 to 360.
Interactive Map is available online at: `http://bit.ly/123TitU` .

detail the price gaps methodology in the following paragraphs.

An excellent point is made in Atkin and Donaldson (2012), that estimating the change in a price gap is not as clear to interpret as would be desirable. They identify three challenges plaguing the price gaps literature. First, product characteristics may be correlated with geography, as with wheat

Figure 2: Data by Year and Connection



13,834 total price observations.

production in Olmstead and Rhode (2002). I attempt to overcome this problem by picking data only for well-defined commodities that were extremely common and homogeneous. Second is that it matters whether or not two locations are physically trading with one another because only then does there exist pressure for the price gap to be reduced to zero. This data is currently not available, but I do have data on some imported goods, and in future drafts it may be possible to trace their exact trade routes for a limited number of locations. The third is that perfect competition may be a strong assumption, and that price gaps are actually equal to trade costs plus markup. My solution to this last problem is to include merchant fixed effects. These fixed effects should account for merchant markups. The downside of this specification is that it implies merchant markups are constant across time and products. Additionally, this number of fixed effects makes standard errors very difficult to compute in the IV specifications [3].

4.1 Estimating Equation

The price gap between merchants c and d at time t takes the following general form:

[3] I am currently unable to compute this IV specification with bootstrapped standard errors using STATA.

$$\log p_{c,d,t} = \beta_0 + \beta_1 \ln dist_{c,d} + \beta_2 connected_{c,d,t} + \alpha X_{c,d,t} + u_{c,d,t} \quad (1)$$

where $X_{c,d,t}$ is a vector of dummy variables to control for geographic variables and data issues, like a source which uses wholesale versus retail prices, or whether a price comes from an Indian Agency. Because the treatment $connected_{c,d,t}$ is highly auto-correlated (a connected city never unconnects and unconnected cities tend to stay unconnected), I follow the suggestion of Bertrand, Duflo and Mullainathan (2004) and use a block-bootstrap method to calculate the standard errors.

The control variables are numerous and can be broken down into two classes: quantitative and indicator. The quantitative control variables are the difference in logs of remoteness, where remoteness for each city is defined as the population-weighted distance to the 10 largest cities in the United States. Great-circle distance between cities and population make up two of the other quantitative controls. The indicator control variables group includes dummy variables for whether one or both locations are an Indian Agency, whether observations are for June 1st or year-average, whether both cities are in the West or if just one city is in the West, whether both cities are on the same river, whether both or one city is on the coast, whether two cities are from the same data source, and whether both reporting merchants are located in the same city. Summary statistics for some of the control variables are given in Table 3.

Previous literature has shown that canals and major waterways are very important to estimating trade costs. For example, Fogel (1964) uses canals as a substitute for railways, Donaldson (2012) uses waterways when calculating lowest-cost trade routes, and Donaldson and Hornbeck (2011) also consider waterways as a major part of their analysis. According to George Rogers Taylor, (1951), every major canal was already in place by 1850, just before my study begins. So I can simply treat a merchant's access to waterways, be those ocean, river, lake, or canal, as a static variable. Since this does not change over time within my sample, I do not have to worry about the confounding effects of waterways, so long as I use appropriate fixed effects or dummy variables for waterways.

4.2 Instrumental Variables

The placing of railroads was not at all a random event. In fact the railroads were the largest infrastructure project attempted in the United States. Clearly a capitalist would only place a line where it would have the largest economic impact. To remedy this selection issue, I use an instrumental variables approach.

In choosing the instrumental variables, I considered the following example: a railroad entrepreneur wants to build a railway line from city c to city d, and does a cost-benefit analysis. Abstracting away from government subsidies and land grants, the benefit to the entrepreneur depends on the amount of potential cargo trade and passenger tickets. These potential sales depend on the population level and the population's characteristics. Secondly, the entrepreneur considers

Table 3: Price Gaps Summary Statistics (236,559 obs)

Variable	Units	Average	Min	Max
Price Gap	¢/pound	7.227	0	2201
Year	years	1872	1851	1892
Distance	kilometers	992.5	0.55	4342.1
Connected	dummy	81.1%	0	1
Both West	dummy	9.4%	0	1
CEW	dummy	27.7%	0	1
Both Coast	dummy	4.7%	0	1
One Coast	dummy	30.7%	0	1
Same River	dummy	6.1%	0	1
Both River	dummy	27.2%	0	1
Manufactured	dummy	24.8%	0	1

the cost of this rail line. The cost of the rail line depends clearly on the distance between the cities, but also on the average difficulty of building each of those miles. So total terrain ruggedness must matter, since rugged terrain is more expensive to develop. Similarly the difference in elevation times the straight-line distance is approximately the proposed track's average grade, and a higher grade will lead to higher building and operating costs. Thus the $connected_{c,d,t}$ indicator variable is endogenous but can be expressed as a function of the sum of the logs of city population, sum of the logs of city ruggedness, and the difference in log elevation times distance [4]:

$$\begin{aligned} connected_{c,d,t} = \alpha_0 + \alpha_1[\ln(Pop_{c,t}) + \ln(Pop_{d,t})] + \alpha_3[\ln(Rugg_c) + \ln(Rugg_d)] \\ + \alpha_4 dist_{c,d} * [\ln(Elev_c) - \ln(Elev_d)] + \eta_{c,d,t} \end{aligned} \quad (2)$$

Each of the potential instrumental variables is plausibly correlated with whether two cities are connected by a railway. Table 4 details the unconditional correlation between each instrument and both the dependent variable, the log of the price gap, and the endogenous variable, connected. Consider in turn the following arguments for each one's exclusion restriction.

Cities have higher costs of living compared to rural areas, which is the effect of agglomeration due to the difficulty of transportation. So in general a city with a high population density is associated with a high price level. However small towns may be geographically isolated which could also lead to higher prices. So it's not clear which would have the higher price, and for which goods. Additionally, consider that one would expect the price gaps between a large and a small city to be large, while the price gap between two small or two large cities to be small. So the dependent variable, the price gap, is not clearly correlated with the sum of the logs of population. My population variable is currently

[4] I also tried using the difference in log rainfall as an insrument but it was too highly correlated with the price gaps to be a good instrument.

not well-defined. It's approximately future population using Census data from 1870, 1880 and 1890. In future drafts I plan to estimate yearly population, and use this as the new instrumental variable.

A rugged city is an expensive city because of transportation costs. Recall the opening quote from C. M. Chase on his visit to Leadville, Colorado in 1881. Leadville lies wedged in a narrow valley between multiple 14,000 foot peaks. Before the railroad was built, prices in Leadville were astronomical, because of the town's isolation, it's booming population, and because the rugged terrain made transportation so expensive. Or consider the infamous Deadwood, South Dakota. Deadwood was an isolated mining town that became famous as one of the most lawless towns in the West, attracting both outlaws and the likes of Wild Bill Hickok and Calamity Jane. Deadwood lies deep in the Black Hills and for its first twelve years all supplies were hauled in by ox-cart and all gold was shipped out by stagecoach. So rugged cities are expensive, but the ruggedness only affects the price gap by determining whether or not a railroad can be built: prices in cities with railroads are not effected by ruggedness.

Differences in elevation have an unclear effect on price gaps. Higher elevation places are typically more remote (ie Leadville), so one would expect a higher price level, but again, price levels are not systematically associated with price gaps. In future drafts I plan to add additional instruments such as distance from a major waterway, as in Keller and Shiue (2008) and distance from a proposed transcontinental route for Western cities, which is similar to the falsification exercise in Donaldson (2012) and the identification strategy in Atack et al. (2009).

Table 4: Correlation of Instrumental Variables

	Dep Var: Log Price Gap	Endog Var: Connected
Sum of Log Ruggedness	0.0286	-0.3107
Log Distance*Difference in Log Elevation	-0.0384	-0.0663
Sum of Log Population	0.0593	0.4633

5 Results

I have results from three different estimation methods. The first two, regressions on price levels and price correlations, are only conditional expectations, and I make no effort at supporting a causal interpretation.

5.1 Price Levels

I noticed an interesting trend when assemblying the data set. It seems that prices for many items were actually *higher* for cities connected to the railroad, and lower for unconnected cities. Railroads were supposed to reduce transportation costs, and theory tells us that when transportation costs are falling, so should the price. Indeed, Column (2) of Table 5 shows that a railroad connection is associated with an increase in the price level of 14 percent. This may at first be unexpected, but I can offer an explanation. At the beginning of my sample, many of the large, commercial cities are already connected, so more rural, distant, or remote cities are all that remain. These remote cities tend to be agricultural. Classical trade theory tells us that when a geographically isolated area goes from autarky to free trade, the price of the good in which the area has a comparative advantage will rise, and the price of the good in which it does not have a comparative advantage will fall. So one would expect more rural areas to have a comparative advantage in agricultural commodities, and a to not have a comparative advantage in manufactured and imported goods, which are made or imported directly into large cities.

Table 5: Regressions on Price Levels

Dep Var: Log(Price)	(1)	(2)	(3)
Connected	0.006	0.142***	0.761***
	(0.0347)	(0.0422)	(0.0312)
Manuf.*Connected	-0.168***	-0.259***	-0.400***
	(0.0444)	(0.0539)	(0.0489)
Connected*West		-0.242***	-0.470***
		(0.0624)	(0.0456)
Connected*West*Manuf.		0.060	0.174*
		(0.106)	(0.0947)
Connected Agency	-0.278***	-0.171	0.044
	(0.0901)	(0.107)	(0.100)
Constant	2.381***	2.319***	1.883***
	(0.0498)	(0.0453)	(0.00733)
Observations	13834	13834	12895
R^2	0.007	0.014	0.002
Number of identifiers	939	939	847
Autoregressive F-stat			35.57

*** p<0.01, ** p<0.05, * p<0.1

Fixed effects are for the year and merchant*commodity.

Columns (1) and (2) have bootstrapped standard errors.

Column (3) uses an AR(1) specification.

Column (1) shows no apparent difference in the price level between connected and unconnected cities for most goods, however connected cities tend to have 17 percent cheaper manufactured goods and connected Indian Agencies tend to have 28 percent overall cheaper goods.

Column (2) of Table 4 is the price level panel regression with bootstrapped standard errors but without an autoregressive term. Being connected to the railroad in the East is associated with a 14 percent overall increase in prices, but with a 26 percent decrease in the price of manufactured goods. On the other hand, western cities see a 24 percent decrease in prices, with no clear difference in the price of manufactured goods.

Column (3) of Table 4 is an AR(1) panel regression on price levels. Connected cities in the East can expect an average increase in agricultural prices of 76 percent and a decrease in the price of manufactured goods of 40 percent. In the West, connected cities are associated with 47 percent decrease in prices but a 17 percent increase in the price of manufactured goods. This last coefficient indicates that the connection for cities in the West is simply less important, or has a lower impact on prices than for cities further east.

Overall, these three regressions show that the correlations are precisely what one would expect: from the theory of comparative advantage we would expect the railroad to bring lower prices for manufactured goods in the West, where they are more scarce, and raise prices for agricultural goods in previously unconnected cities where agricultural goods were relatively abundant.

These results are consistent with previous work. Atack, Haines and Margo (2008) find that railroad access effectively increased the size of markets available to producers which allowed them to grow in size. My findings show that when cities connect through the railway network, the lower prices for manufactured goods in the East greatly influence the price of the newly connected city, indicating that market access was in fact important for producers of manufactured goods. Similarly, Gutberlet (2013) who found that railroad access encouraged concentration of industries in Germany in this same time period. This is consistent with the price level story because the theory of comparative advantage tells us that manufacturing industries will concentrate where costs are lowest.

Additionally, these results are also consistent with the findings in Atack and Margo (2011) which found that railroads induced settlers in the Midwest to invest heavily in private capital by developing farmland. Again, the connection variable has a positive sign for agricultural goods, likely indicating that farmers are receiving more money for their products. This in turn would encourage residents or settlers to develop land which is more marginal and further from loci of demand.

5.2 Price Correlations

Based on the methodology of Engel and Rogers (1996), my second exercise is to calculate the change in price correlations before versus after two cities are connected through the railroad network. I define these price correlations as the standard deviation of the difference in prices between cities c and d over the pre-rail period (if that period exists for the city-pair) with the $connected_{c,d,t}$ indicator variable equal to zero. Over the rail-connected period I change the $connected_{c,d,t}$ indicator variable to one. When calculated in this way, about 75 percent of the sample is connected to the railroad.

Table 6: Regressions on Price Correlations

Dep Var: Log(StdDev(Price Gap))	(1)	(2)
Connected	-0.0804***	0.192***
	(0.0306)	(0.0370)
Log(Dist)	0.0377***	0.0352***
	(0.00544)	(0.00568)
Cross East-West		0.0167
		(0.312)
Cross East-West*Connected		-0.392***
		(0.0476)
Cross East-West*Connected*Manuf		-0.0763*
		(0.0428)
Both West		-0.101
		(0.615)
Both West*Connected		-0.595***
		(0.0690)
Both West*Connected*Manuf		-0.134***
		(0.0422)
Never Connected		-0.294***
		(0.0662)
Constant	-1.067***	-1.000***
	(0.220)	(0.236)
Observations[1]	224959	224959
R^2	0.731	0.737

[1] 224,959 observations consisting of 25,913 groups; regressions run at the group (city-pair at time t) level, where $t = 0$ or 1 for before or after connected to the railroad, and regressions weighted by number of observations per group.
Robust standard errors in parentheses, clustered by city-pairs.
*** p<0.01, ** p<0.05, * p<0.1
Fixed Effects are for both commodities and individual cities.
Other variables included in both regressions include controls for source data and types of goods, as well as geographic variables.

Table 5 has the results from two regressions on price correlations. Column (1) indicates, contrary to theory, that prices in connected cities covary negatively, albeit an economically small amount of 8 percent. The regression in Column (2) breaks the city-pairs into groups of (i) both east, (2) one east and one west, and (3) both west, and shows the interactions between region, manufactured goods, and connected status. The values themselves aren't terribly informative, but the idea to take away from here is that it's very important to distinguish between city-pair locations, manufactured versus agricultural goods, and connected versus unconnected. The cities in the West have prices which correlate very different from eachother and from cities in the East.

I don't go into additional detail here because the problem is that this specification fails a falsification exercise, which is not shown. I artificially change the year that a city connects to the network, and the results remain statisically sigificant. This may be because the difference in price correlations from connected to unconnected is so strong that changing a single year still results in statistically significant results. So while this price correlation exercise may be informative, I cannot rule out spurious correlation. This brings us to the last exercise: estimation using price gaps.

5.3 Price Gaps

This last section of results is the truly informative one. Table 7 shows the result of the regressions discussed in the Empirical Implementation section.

Column (1) of Table 6 gives the OLS results and column (2) of Table 6 gives the full-specification IV results. The main result is that connected cities have a 70 percent lower price gap than unconnected cities. A variety of full IV regression specifications (not shown) put this number between negative 70 and negative 80 percent. I am further convinced this is the correct value based on an estimate of negative 73 percent in an earlier version of Donaldson's (2012) "Railroads of the Raj" paper. Columns (3) through (5) give the results for one instrument at a time, which are mostly shown for verification of the F statistic.

The equally other important result here in the division of manufactured versus agricultural goods and East versus West. Overall, connected cities have a higher cost for manufactured goods. But the price gaps for one city in the East and one in the West is 6.6 percent lower, and if both cities are in the west, then they have a 4.6 percent lower gap. One would expect price levels to change in heterogeneous ways, as my previous results have shown. However these price gaps show considerable heterogeneity based on where a merchant is located in the country and on whether the good is manufactured or agricultural.

5.4 External Validity

The estimates of the affect of railway access that I have examined are unique to the United States and the time period. These only indicate a lower bound of the actual effect of railroads on prices for the following reasons. First of all, the American West was sparsely settled by Native Americans around the beginning of this time period. Retail establishments did not widely appear until after a major wagon route or railroad was established. So a merchant that is established after the railroad comes to town usually has no precedent with which to compare. Second of all, dozens, if not hundreds of railroad companies operated over the railroad network. Unlike today, trains at the time

Table 7: Regressions on Price Gaps

Dep Var: Log(Price Gap)	(1)	(2)	(3)	(4)	(5)
Estimation Method	OLS	IV	IV	IV	IV
Connected	-0.0846***	-0.700***	-0.339***	-1.709***	-0.844***
	(0.0169)	(0.0750)	(0.0970)	(0.401)	(0.0816)
Connected*Manuf	-0.0901**	0.344***	0.0890**	1.054***	0.445***
	(0.0343)	(0.0682)	(0.0717)	(0.296)	(0.0811)
East-West*Connected*Manuf	-0.125***	-0.0658***	-0.101***	0.0316	-0.0520**
	(0.0411)	(0.0459)	(0.0431)	(0.0734)	(0.0433)
Both West*Connected	-0.0998***	0.330***	0.0779**	1.035***	0.431***
	(0.0374)	(0.0627)	(0.0748)	(0.272)	(0.0713)
Both West*Connected* Manuf	-0.00281	-0.0463*	-0.0208	-0.118***	-0.0564**
	(0.0520)	(0.0555)	(0.0530)	(0.0660)	(0.0544)
Constant	-0.251***	0.0125	-0.142***	0.444**	0.0739
	(0.0401)	(0.0501)	(0.0550)	(0.185)	(0.0515)
Observations	236559	236559	236559	236559	236559
R^2	0.394	0.384	0.392	0.327	0.379
Instrumental Variables:					
Ruggedness		yes	yes		
Dist*Diff Elevation		yes		yes	
Population		yes			yes
First-stage F statistic		5926.8	8165.2	1096.1	10421.7

Bootstrapped standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Fixed Effects for cross-regions: there are 8 regions, making 35 cross-region fixed effects.

Column (1) is the full sample with an OLS specification.

Columns (2) through (5) are IV specifications. Column (2) is the preferred, full IV specification.

never ran on the tracks of a competitor. And even if they did, the differences between rail gauges meant that they wouldn't get far. So hopefully if a railroad network were built from scratch today, it could follow a better, more integrated model. [5]. Third, technology and safety were improving at an impressive rate during this period. Since that time they have continued to improve, but the marginal improvement remains unprecedented. And fourth, corporations and railroads were as inseparable from one another as either were to corruption. Corruption presumably isn't as much of a problem anymore and hopefully corporate governance has improved greatly since that time. So considering all of these reasons why the time period and subject is unique, I would fully expect the effect of railroads on prices to be larger in other instances.

[5] The standard of four feet, eight and a half inches was not fully implemented in my sample according to White (2011, pg 3).

6 Conclusion

Few scholars have ever questioned that railroads lead to tremendous growth in the Western United States in the 19th Century, but growth of the railroads and growth of economic activity are so heavily intertwined that it becomes difficult to identify the effects of the railroads. I have implemented an instrumental variables approach on commodity price gaps in an attempt to identify these effects. I have also shown that railroads affect cities in the west in more dramatic ways: lowering the price gap, but this is actually associated with a price increase for agricultural goods, and a price decrease for manufactured goods. This paper sheds light on the effect of railroads in decreasing price dispersion in the United States in the 19th Century. I have barely scratched the surface of available data for the United States in this time period, though I have compiled the most high-quality data. Future drafts could include data which is older and more geographically diverse: such as from the Pacific Northwest or more from the South. Also expect better data on population, including estimates by year, and more detailed analysis can be done with imported goods. In fact with goods such as Green Rio coffee and Puerto Rican molasses, I can, as in Atkin and Donaldson (2012), trace probable trade routes and more accurately estimate exact transportation costs.

References

[1] Aldrich, Nelson. (1893) *Wholesale Prices, Wages, and Transportation: Report from Mr. Aldrich, from the Committee on Finance, March 3, 1893.* Committee on Finance. Senate. 52nd Congress. 1st Session. Report 986. Parts 1-4. Washington, DC: U.S. Government Printing Office.

[2] Atack, Jeremy, Matthew Jaremski, and Peter L. Rousseau. (2013) "American Banking and the Transportation Revolution Before the Civil War" *working paper.*

[3] Atack, Jeremy and Robert Margo. (2011) "The impact of access to rail transportation on agricultural improvement: The American Midwest as a test case, 1850-1860" *The Journal of Transport and Land Use* 4(2) (Summer 2011): 5-18.

[4] Atack, Jeremy, Fred Bateman, Michael Haines, and Robert Margo. (2009) "Did Railroads induce or follow economic growth? Urbanization and population growth in the American Midwest, 1850-60." *NBER Working Paper no. 14640*, January 2009.

[5] Atack, Jeremy, Michael Haines and Robert Margo. (2008) "Railroads and the Rise of the Factory: Evidence for the United States, 1850-70," chapter in *Economic Evolution and Revolution in Historical Time* 162-179.

[6] Atkin, David and Dave Donaldson (2012) "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs", working paper.

[7] Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan. (2004) "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics* 119(1): 249-275.

[8] Chase, C. M. (1882) *New Mexico and Colorado in 1881* Frontier Book Co., Fort Davis, Texas. 1968 edition.

[9] Chen, Yongmin and Michael Riordan. (2007) "Price and Variety in the Spokes Model", *The Economic Journal* 117(July): 897-921.

[10] Coelho, Philip and James Shepherd. (1974) "Differences in Regional Prices: The United States, 1851-1880", *The Journal of Economic History* 34(3): 551-591.

[11] Donaldson, Dave and Richard Hornbeck. (2013) "Railroads and American Economic Growth: A 'Market Acess' Approach", NBER working paper no. 19213, July 2013.

[12] Donaldson, David. (2012) "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure" NBER working paper no. 16487, October 2010. Forthcoming in *American Economic Review.*

[13] Engel, Charles and John Rogers. (1996) "How Wide is the Border?" *American Economic Review* 86(5): 1112-1125.

[14] Fishlow, Albert. (1966) "Productivity and Technological Change in the Railroad Sector, 1840-1910" in *Output, Employment, an Productivity in the United States after 1800* by Dorothy S. Brady (ed.), NBER, pages 583-646.

[15] Fogel, Robert. (1964) *Railroads and American Economic Growth: Essays in Econometric History* Johns Hopkins Press: Baltimore, Maryland.

[16] Gutberlet, Theresa (2013) "Railroads and the Regional Concentration of Industry in Germany 1846 to 1882" *working paper*

[17] Hoover, Ethel. (1958) "Wholesale and Retail Prices in the Ninteenth Century", *Journal of Economic History* 18(3): 298-316 (Sep. 1958).

[18] Hoover, Ethel. (1960) "Retail Prices After 1850", a chapter in *Trends in the American Economy in the Nineteenth Century*, from the NBER, ISBN 0-870-14180-5. pages 141-190.

[19] Jacks, David. (2006) "What drove 19th century commodity market integration?" *Explorations in Economic History* 43 (2006): 383-412.

[20] Keller, Wolfgang and Carol Shuie. (2008) "Institutions, Technology, and Trade" *University of Colorado Working Paper*, December 2008.

[21] Munshi, Kaivan. (2003) "Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market" *Quartely Journal of Economics* 118(2): 549-599.

[22] Olmstead, Alan and Paul Rhode. (2002) "The Red Queen and the Hard Reds: Productivity Growth in American Wheat, 1800-1940" *The Journal of Economic History* 62(4): 929-966.

[23] Santos Silva, Joao and Silvana Tenreyro. (2005) "The log of gravity" CEP Discussion Paper No. 701 (July 2005).

[24] Statistical Abstract of the United States 1900 Twenty-Third Number. (1901)

[25] Taylor, George Rogers. (1951) *The Transportation Revolution.* Rinehart, New York.

[26] Weeks, Jos. D. (1883) "Report on the Average Retail Prices of the Necessaries of Life in the United States". Supplement to the *1880 Census of the United States.*

[27] White, Richard (2011) *Railroaded: The Transcontinentals and the Making of Modern America* W. W. Norton and Company, 1st Edition, May 2011.

A Data Appendix

This Appendix provides a detailed description of the paper's hand-collected dataset.

First, the commodity prices came from three sources[6]. Prices in San Fransisco are from Berry (1984). He provides a table of the average annual prices of 34 commodities from 1847-1900 on pages 241-243. Aldrich (1893) gives spotty data from the War Department for 20 commodities in 11 cities on pages 1652-1810, for the years 1840-1891, excluding 1861-1865. The data series vary between yearly, quarterly, and monthly observations. I find the average price for the year. The Aldrich Report also has yearly prices paid by the Navy from 1840-1892 for 22 commodities primarily in New York, Boston, and Norfolk, but also at 11 other naval bases for fresh beef. Whenever two different prices are given for an observation, as is frequently the case in the Aldrich data, I take the average of the two observed prices. Lastly, the Aldrich Report gives time series of the prices of commodites in Prussia, Italy, Austria, France, and England, and for the cities of London and Hamburg. For this draft of the paper I have chosen not to include these European cities. My final and largest source for price data is the Weeks Report (1883), an addendum of the 1880 Census. Pages 3-93 are tabulations of questionnaires sent to retailers, asking them about the price of their goods from 1851 up to the collection date in 1880. The report contains yearly prices for approximately 49 goods, with most of the 95 or more retailers selling less than half of those 49 goods. The prices for Boston are the range of prices found in the "The New England Farmer". I use the midpoint of that range. The Weeks Report provides over 70 percent of the observations, but the Berry book and the Aldrich Report provide relatively more geographic variation, and a higher proportion of observations from cities not connected to the railroad. These are the only comprehensive data sources that I have been able to locate. I do have a few other series not yet included, but these observations easily number under 100. I have found quite a bit of data from digitized newspapers, but because of the number of newspapers and their frequency, extracting the data in a systematic way is a daunting task.

Data on the year that a city connected to the railroad was more difficult to ascertain. Once the cities of interest were tabulated, I checked the location of each against a map of all U.S. railroads in 1851. I chose 1851 because it is the starting year for the Weeks dataset. For those cities not connected in 1851, I use a variety of sources. For Key West, I looked in Standiford (2002). For the other two Florida cities, Pensacola and Jacksonville, I read the text in Turner (2008). I likewise read the text in Hayes (2010) for the connection years for St. Paul and Cedar Rapids. I consulted maps in Hayes (2007) to find the year that Mare Island (Vallejo) California was connected to San Francisco and the rest of the network. For the remaining cities, I consulted the David Rumsey map collection for each year between 1851 and 1892. I tend to prefer the text explanations of railroad connections over the map images because cartographers tended to make mistakes with the length and route of railroads. These accidents occurred because railroads at this time inherently grew at varying paces, and keeping up with each of the hundreds of railway companies in any given year would have been difficult. For the Indian agencies, the price data begins in 1877 and ends in 1892. Because of this, I started looking at a map of railroads in 1876. If an Indian agency was connected on or before 1876, I simply coded that location as being connected in 1876.

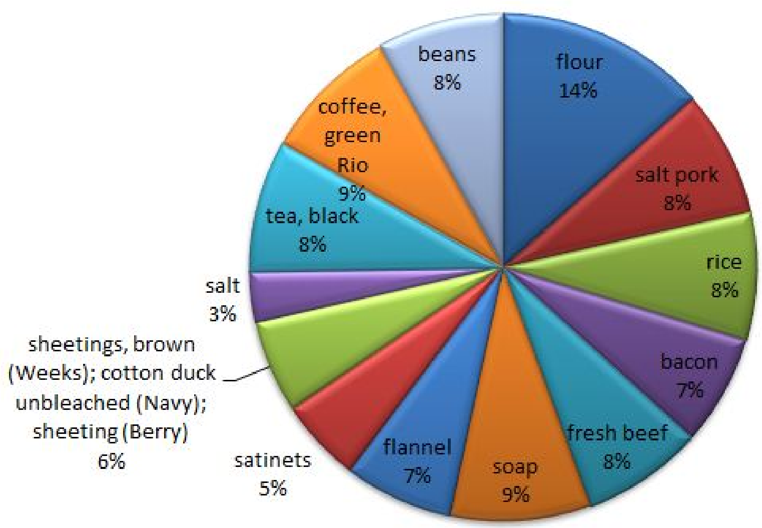
The rest of the data points were not particulary difficult to ascertain. Geographic coordinates, used to compute great-circle distance between pairs of cites, comes from ArcGIS. City population data, one of the instrumental variable, come from the 1870, 1880, and 1890 U.S. Censuses. I have imputed all other data, such as the dummy variables for "West", "coast", and for several major navigable waterway (Mississippi, Missouri, Hudson/Erie Canal, Great Lakes, Delaware, and Ohio).

[6] The Bureau of Labor Statistics did not begin collecting price data by city till 1913.

Just a few additional notes on coordinates. Some of the locations of each of the 157 Indian Agencies were difficult to obtain. Most of the Indian agencies became incorporated towns of the same name, so that their locations were stored in ArcGIS. Some of the locations, such as Cale, Kansas are significant enough that, despite no longer having a physical presence, are still stored as historical locations in map databases such as Google. Most of the agencies can be found on the family search website [7] which describes the location of many of the Indian agencies, most of which I later checked with historical maps. I was unable to locate four agencies, each of which only had one data point each. Lastly, Sidney was labeled as being located in Wyoming, but I could find no evidence of such a place, so I set the coordinates as Sidney, Nebraska, just 56 miles from the Wyoming border, where a regional military outpost was known to have existed.

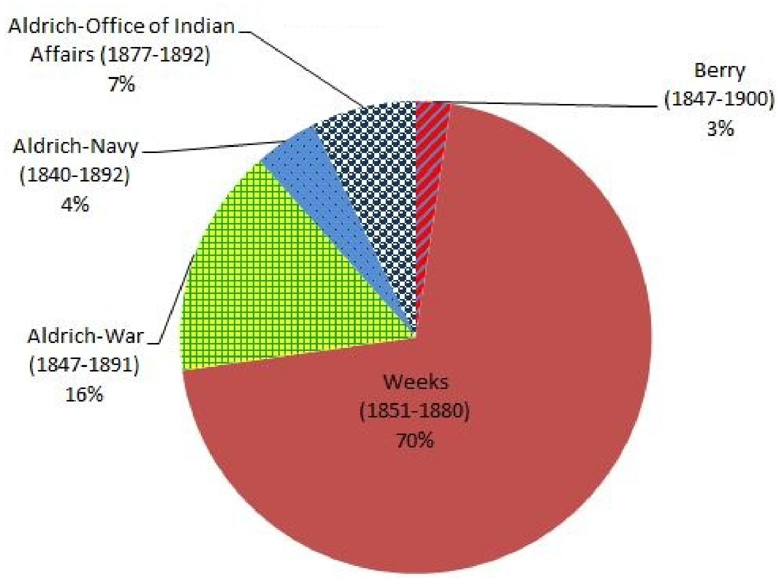
Both the elevation and precipitation data are available free online in R libraries. I have posted the code to extract this data on my website. Ruggedness is calculated as in Riley et al (1999) and Nunn and Puga (2012) from data at the 30 arc second (aka 1km) resolution. This data is known as SRTM30, and originally came from the National Imagery and Mapping Agency of NASA. Data was downloaded from http://diva-gis.org/Data. Precipitation data was taken from the worldclim database (Hijmans et al. 2005). This dataset uses precipitation data from 47,554 locations. The interpolated data is available at the 30 arc second (1km) spatial resolution, and I used the 10 arc minute resolution. Values are 60-year averages. Available for download from: http://www.worldclim.org/.

Figure 3: Data by Commodity Type



[7] https://www.familysearch.org/learn/wiki/en/American_Indian_Genealogy

Figure 4: Data by Source



References

[1] Aldrich, Nelson. (1893) *Wholesale Prices, Wages, and Transportation: Report from Mr. Aldrich, from the Committee on Finance, March 3, 1893*. Committee on Finance. Senate. 52nd Congress. 1st Session. Report 986. Parts 1-4. Washington, DC: U.S. Government Printing Office.

[2] Berry, Thomas Senior. (1943) *Western Prices Before 1861: a Study of the Cincinnati Market*. Harvard University Press, Cambridge, MA.

[3] Berry, Thomas Senior. (1984). *Early California: Gold Prices Trade* The Bostwick Press, Richmond, VA. (HB235.U6 B37 1984).

[4] Coelho, Philip and James Shepherd. (1974) "Differences in Regional Prices: The United States, 1851-1880", *The Journal of Economic History* 34(3): 551-591.

[5] Cole, Arthur Harrison. (1938) *Wholesale Commodity Prices in the United States, 1700-1861* Harvard University Press, Cambridge, MA.

[6] Hayes, Derek. (2007) *Historical Atlas of California* University of California Press.

[7] Hayes, Derek. (2010) *Historical Atlas of the North American Railroad* University of California Press, Berkely and Los Angeles, CA.

[8] Hijmans, Robert, Susan Cameron, Juan Parra, Peter Jones and Andy Jarvis. (2005). "Very high resolution interpolated climate surfaces for global land areas" *International Journal of Climatology* 25: 1965-1978.

[9] Modelski, Andrew. (1984) *Railroad Maps of North America: The First Hundred Years.* Library of Congress, Washington, D.C. ISBN 0844403962.

[10] Nunn, Nathan and Diego Puga. (2012) "Ruggedness: The Blessing of Bad Geography in Africa", *The Review of Economics and Statistics* 94(1): 20-36.

[11] Riley, Shawn J, Stephen DeGloria, and Robert Elliot. (1999) "A terrain ruggedness index that quantifies topographic heterogeneity". *Intermountain Journal of Sciences* 5(1-4): 23-27.

[12] Standiford, Les. (2002) *Last Train to Paradise.* Crown Publishers, New York. 1st Edition.

[13] Taylor, George Rogers. (1932) "Wholesale Commodity Prices at Charleston, South Carolina, 1796-1861" *Journal of Economic and Business History* 4(August): 848-868.

[14] Turner, Greg. (2008) *A Journey into Florida Railroad History.* University of Florida Press, Gainesville, FL.

[15] U.S. Bureau of the Census. *The Statistics of the Population of the United States: Ninth Census (1870)* Volume I, Table III: Population of Civil Divisions Less Than Counties, 77-296.

[16] U.S. Bureau of the Census. *1880 Census of the United States* Part 4: Report on the Agencies of Transportation in he United States Including the Statistics of Railroads, Steam Navigation, Canals, Telegraphs, and Telephones, 296.

[17] U.S. Bureau of the Census. *Statistics of the Population of the United States at the Tenth Census (1880)* Part 1. Table III: Population of civil divisions less than counties, 87-375.

[18] U.S. Bureau of the Census. *Report on Population of the United States Eleventh Census: 1890* Part 1. Table 5: Population of states and territories by minor civil divisions, 52-308.

[19] U.S. Bureau of the Census. *Statistical Abstract of the United States. 1900. Twenty-Third Number.* Railroads: Tables 100-111, 374-384.

[20] Warren, G.F. and F.A. Pearson (1939) "Index of Wholesale Prices, Variable Group Weights", in *Prices* John Wiley, New York, New York: 12-13. (Data downloaded from the NBER Macrohistory Database IV: Prices.)

[21] Weeks, Jos. D. (1883) "Report on the Average Retail Prices of the Necessaries of Life in the United States". Supplement to the *1880 Census of the United States.*