How much did Railroads Affect Market Integration? Evidence from 19th Century France

Preston Mui
University of California, Berkeley

May 12, 2016

Abstract

Abstract text [1] [2]

[1] I thank Sam Leone, Isabela Manelici, Patrick Russo, Brad Delong, and Barry Eichengreen for useful comments and discussion. The usual disclaimer applies. All mistakes are my responsibility.

[2] All data and code for this project is available at https://github.com/PrestonMui/frenchrailroads.

1 Introduction

2 Data

rain prices, measured in centimes per hectoliter, were collected biweekly by local officials in various French *communes* between 1825 and 1914, and sent to the national government. These data now reside at the French National Archives, and the price data from 53 *communes* are publicly available from Roehner (2016).[3] The sample of *communes* from Roehner (2016) cover nearly all geographic areas of France, with the exception of Alsace-Lorraine near the border of Luxembourg. In general, the sample tends to include larger cities and port cities, as well as Northern cities due to higher wheat production in the North. Figure 1 shows the location of the *communes* comprising the price data.

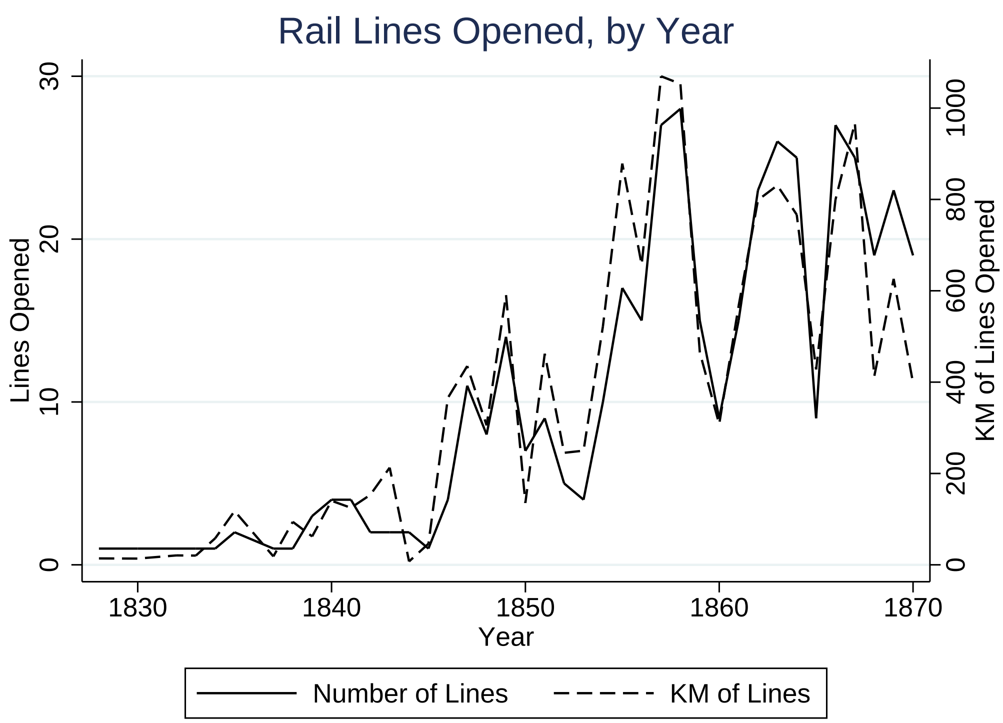
Figure 1: Location of *commune* in price data (Roehner, 2016).

I obtain information on railroad openings from Palau (1998, 2001, 2004), who uses primary sources from the Musee Francais du Chemin de Fer (The French Train Museum), SNCF archives, and *Annales de Mines*, a French technical mining journal from the 1800s. These sources contain information on every railroad opening

[3]Some *communes* were eliminated by Roehner (2016) due to imprecise or incomplete data.

in France between 1828 and 1870, including the location of the two endpoints, the day the route was opened, and the distance of the route. Overall, there were 435 rail lines opened in France between 1828 and 1870. Figure 2 shows the number and total distances of railroad lines opened in each year during the sample. To get geographical distance information, I feed both the *communes* from the grain price data and the rail stations into the Google Maps Geocoding API (Google Developers, 2016) for latitudinal and longitudinal coordinates, and apply the haversine formula to calculate as-the-crow-flies distance between each pair of rail stations and *communes*.

Figure 2: Rail openings over time, France, 1825 - 1870 (Roehner, 2016).

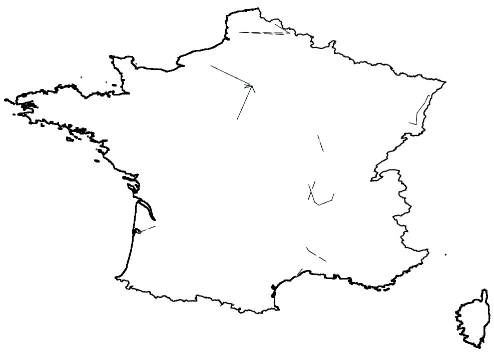


From 1828 to 1870, there were a total of 410 routes opened, with a total cumulative distance of 15,445 kilometers[4]. The average route was 30 kilometers long, with the largest being the 203 kilometer Paris-Lille line, which opened in June 1846. Figure 3 shows the evolution of rails across the country. Each map shows the railroads opened by the end of the year listed.

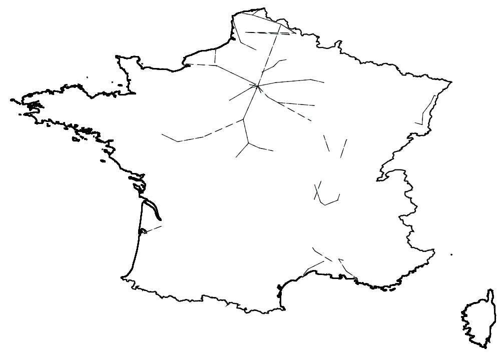
Some matching betweeen the rail openings data and the grain data is required, as our interest is of rail connections between price locations, not rail stations. To this end, at any particular point in time during the sample period, I consider two

[4]Distances are as-the-crow flies, not actual distances accounting for curves.

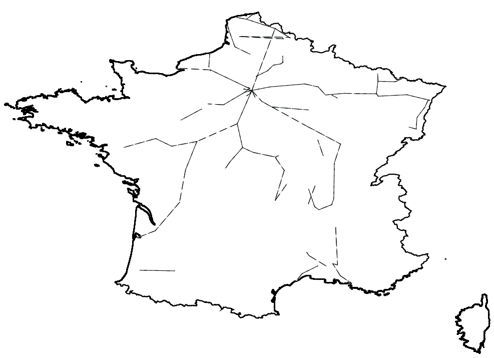
(a) 1845. 27 routes, 1,104 km.



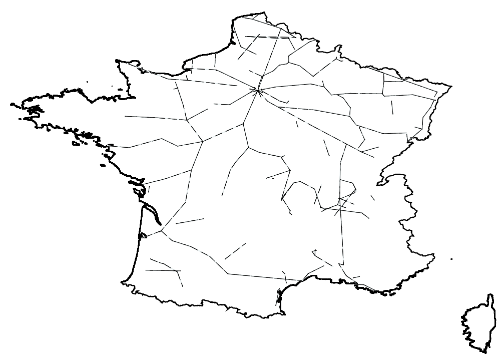
(b) 1850. 71 routes, 2,937 km.



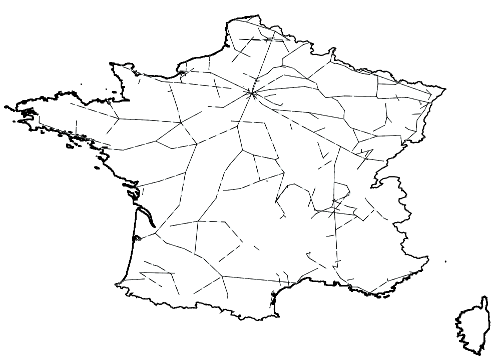
(c) 1855. 114 routes, 5,296 km.



(d) 1860. 207 routes, 8,849 km.



(e) 1865. 302 routes, 12,240 km.



(f) 1870. 410 routes, 15,445 km.

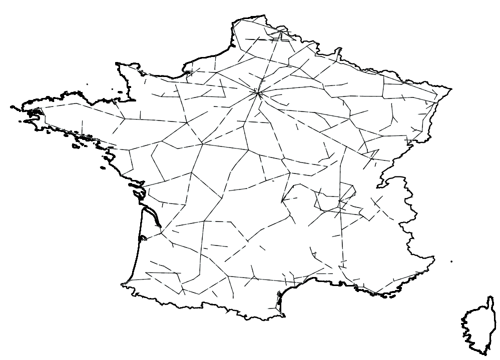


Figure 3: Railroads open at year-end. Endpoints of lines are stations, but the actual lines are simply as-the-crow-flies lines between stations, not actual rail routes.

rail stations to be connected if they are connected in a graph with rail stations as vertices and railroads as edges. Then, I consider two *communes* to be "connected" by railroads at a particular point in time if any rail station within 40 kilometers of the first *commune* is connected to any rail station within 25 kilometers of the second *commune*. I assume that connections are symmetric in the sense that connections are unaffected by the direction of travel; that is, the statement "Soissons is connected to Albi" is equivalent to "Albi is connected to Soissons."

By 1870, 41 out of 53 *communes* in the grain price data are associated with some rail station by this metric.[5]

3 Estimation

Ideally, one would like to would like to estimate something similar to a vector error-correction model (VECM), as in Johansen and Juselius (1990). This model would estimate equilibrium multilateral relationships between price series in multiple cities and the the speed at which the time series converge to this equilibrium. However, using this approach is somewhat unworkable with fifty-three price series, because the number of observations (across time) required to estimate this model would make it difficult to link changes in the relationship between price series to the introduction of railroads. Furthermore, the complexity of this model would make it difficult to interpret the estimated coefficients of this model.

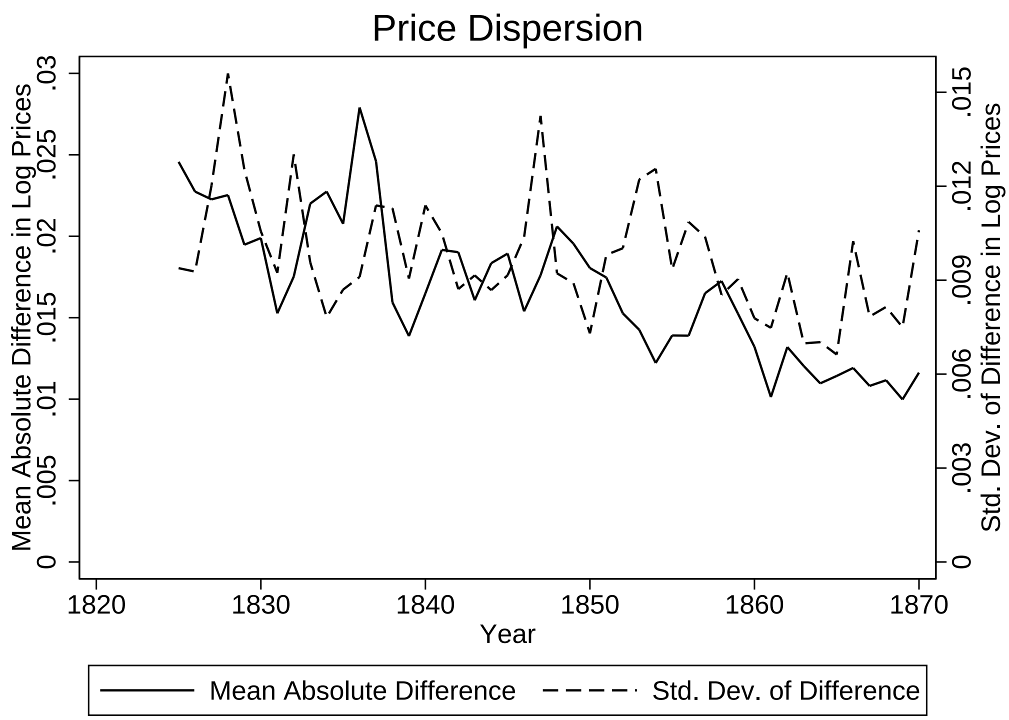
Instead, I take the simpler approach of examining the effect of rail connections on both the magnitude and volatility of price differentials between *communes*. To examine the magnitude of price differentials, I construct $P_{i,j,t}$ as the mean of the absolute value of differences in logged prices in *communes* i and j in year t. I measure the volatility of the price differential, $V_{i,j,t}$, as the standard deviation of differences in logged prices. Although this is less precise than estimating the dynamic relationship between price series in a VECM, linking changes in these variables to rail connections is much easier. I plot the annual averages, across all *commune* pairs, of $P_{i,j,t}$ and $V_{i,j,t}$ in Figure 4. Price dispersion, as measured by absolute price differentials, fell over the sample period from 2.5% to 1%. The volatility of price differentials appears to have fallen over time as well (I estimate a coefficient of -0.0006 on year on a best-fit line, significant at the 1% level), but the pattern is much less pronounced than the fall in absolute differentials.

I begin by estimating the following regression:

$$P_{i,j,t} = \beta_0 + \beta_d d_{i,j} + \beta_R R_{i,j,t} + \beta_{Rd}(R_{i,j,t} \times d_{i,j}) + \sum_{k=1}^{N} \gamma_k C_k + \sum_{s=1}^{T} \eta_s Y_s + \epsilon_{i,j,t} \tag{1}$$

[5] The *commune* that are not associated with a rail station are Albi, Annecy, Arras, Bayeux, Blois, Carcassonne, Dieppe, Digne, Macon, Peyrehorade, Pontlabbe, Vannes

Figure 4: Price Dispersion over Time



where $d_{i,j}$ is the log distance between *commune* i, j; $R_{i,j,t}$ is a dummy equal to one if i and j are connected by railroads, as defined above. C_k is a dummy equal to 1 if *commune* k is one of the *communes* in the pairwise observation (so each *commune* pair has two instances of C_k equal to one), and the Y_s terms are year fixed-effects. This regression takes the form of a difference-in-difference regression, with the existence of a rail connection as the treatment, and the with the effect of a rail connection varying by distance between the *communes*.

I present results from (1) in Column 1 of Table 1. As expected, the coefficient on log distance is positive and significant, indicating that price dispersion increased with distance. What is less expected is the significant positive coefficient on the rail dummy, which suggests that railroad connections *increased* the degree of price dispersion between cities. However, the significant and negative coefficient on the interaction of the rail dummy and log distance suggests that the presence of railroads attenuated the effects of distance on price dispersion. Because the distance between *commune* is generally fairly large, the combination of the rail dummy and its interaction with log distance implies that railroads lowered price dispersion for *commune* pairs that were at least 140 kilometers apart, which accounts for over 85% of the *commune* pairs for which there was a rail connection at the end of the sample.

Table 1: Dependent Variable: Average Absolute Log Price Difference

	(1) Baseline	(2) Min Dist 100KM	(3) Min Dist 150KM	(4) Min Dist 200KM
Log Dist. (km)	0.00385***	0.00368***	0.00366***	0.00366***
	(0.00008)	(0.00007)	(0.00007)	(0.00007)
Rail Dummy	0.00792***			
	(0.00068)			
Rail Dummy x Log Dist. (km)	-0.00160***			
	(0.00012)			
Rail Dummy x (Dist ¿ Dmin)		0.00142***	0.00028	-0.00075***
		(0.00023)	(0.00021)	(0.00021)
Rail Dummy x Log (Dist. - Dmin)		-0.00229***	-0.00214***	-0.00168***
		(0.00017)	(0.00021)	(0.00027)
Constant	-0.00235*	-0.00139	-0.00131	-0.00132
	(0.00114)	(0.00112)	(0.00112)	(0.00112)
Observations	59527	59527	59527	59527
Adjusted R^2	0.329	0.329	0.329	0.329

All regressions use Year and Commune Fixed Effects

One possible explanation of this result is that rail might not significantly reduce trade costs between *commune* that are very close to each other, and have no effect on price dispersion. With this in mind, I estimate the following specification:

$$P_{i,j,t} = \beta_0 + \beta_d d_{i,j} + \beta_{\tilde{R}} \tilde{R}_{i,j,t} + \beta_{\tilde{R}\tilde{d}} (\tilde{R}_{i,j,t} \times \tilde{d}_{i,j}) + \sum_{k=1}^{N} \gamma_k C_k + \sum_{s=1}^{T} \eta_s Y_s + \epsilon_{i,j,t} \quad (2)$$

where $\tilde{d_{i,j}} = \max\{\log(d_{i,j}) - \log(d_{min}), 0\}$ for some minimum distance d_{min}. This specification constrains the effect of rail connections on price dispersion to be zero when the *communes* are closer than d_{min} kilometers, attenuating the effects of distance thereafter. I present the estimates of (2) using $d_{min} = 100, 150$, and 200 kilometers in columns (2) - (4) in Table 1.

I present results from estimating (1) and (2) with $V_{i,j,t}$ as the dependent variable are presented in Table 2. As with the results on absolute differential, the coefficients on the rail dummy interactions with the distance variables are significant and negative, suggesting that railroads attenuated the effects of distance on price dispersion. Again considering a hypothetical "average" commune pair separated by 345 kilometers, the total estimated effect of a railroad is to attenuate the total effect of distance on volatility by between 4.6% and 5.6%. For the most-distant *commune* pair (Pont-l'Abbe and Digne), the attenuation is between 7.7% and 8.7%.

Table 2: Dependent Variable: Std. Deviation of Absolute Log Price Difference

	(1) Baseline	(2) Min Dist 100KM	(3) Min Dist 150KM	(4) Min Dist 200KM
Log Dist. (km)	0.00221***	0.00211***	0.00211***	0.00211***
	(0.00003)	(0.00003)	(0.00003)	(0.00003)
Rail Dummy	0.00354***			
	(0.00029)			
Rail Dummy x Log Dist. (km)	-0.00071***			
	(0.00005)			
Rail Dummy x (Dist ¿ Dmin)		0.00011	-0.00030**	-0.00046***
		(0.00011)	(0.00010)	(0.00009)
Rail Dummy x Log (Dist. - Dmin)		-0.00062***	-0.00048***	-0.00043***
		(0.00008)	(0.00010)	(0.00012)
Constant	-0.00503***	-0.00448***	-0.00447***	-0.00447***
	(0.00077)	(0.00077)	(0.00077)	(0.00077)
Observations	59527	59527	59527	59527
Adjusted R^2	0.296	0.296	0.296	0.296

All regressions use Year and Commune Fixed Effects

Although the results are statistically significant, are they economically significant? To answer this, I consider the effects of rail presence on two hypothetical *commune* pairs, one separated by 350 kilometers, and the other separated by 900 kilometers. I chose these values because 350 kilometers was approximately the mean distance of rail-connected *communes* at the end of the sample period, and 900 kilometers is the maximal distance of any two rail-connected communes (Pont l'Abbe, in the Northwest, and Digne, in the Southeast). Then, using the estimated coefficients from the preceding regressions, I calculate an estimated "treatment effect" for these two hypothetical *communes*, and present the values in Table 3.

Table 3: Implied rail effects on hypothetical *communes*: price differentials

	"Average" *Commune*		"Maximal" *Commune*	
Specification	Change in $P_{i,j,t}$	Attenuation of Distance Effect	Change in $P_{i,j,t}$	Attenuation of Distance Effect
Baseline	-.0014266	6.4%	-.0029599	11.3%
$d_{min} = 100$	-.001412	6.6%	-.0036037	14.4%
$d_{min} = 150$	-.0015041	7.0%	-.0035527	14.3%
$d_{min} = 200$	-.0016673	7.8%	-.0032794	13.2%

This back-of-the-envelope suggests the effect, though statistically significant, is small. The implied estimated effect of a rail connection on the "average" *commune* pair was to reduce the mean logged price difference by between 0.14% and 0.016%, depending on the specification used. To understand the economic significance of this reduction, I compare this effect to the implied effect of distance ($\beta_d * d_{i,j}$) on price dispersion to calculate the amount of price dispersion caused by distance that is "attenuated" by the rail connection. For this "average" commune, the effect is between 6.4% and 7.8%, depending on the specification used. Even for the two rail-connected *communes* furthest apart (Pont-l'Abbe, in the Northwest, and Digne, in the Southeast, were 900 kilometers apart), the estimated attenuation effect was less than 15% across all specifications.

The results for price differential volatility are similar. I calculate the implied effects of rail connections on the same hypothetical *commune* pairs and present the results in Table 4. As with the results on absolute levels of price dispersion, the presence of a rail connection between the hypothetical *communes* only attenuates a small fraction of the distance effect on price difference volatility. This result holds across the different specifications for both the "average" and "maximal" hypothetical *commune*.

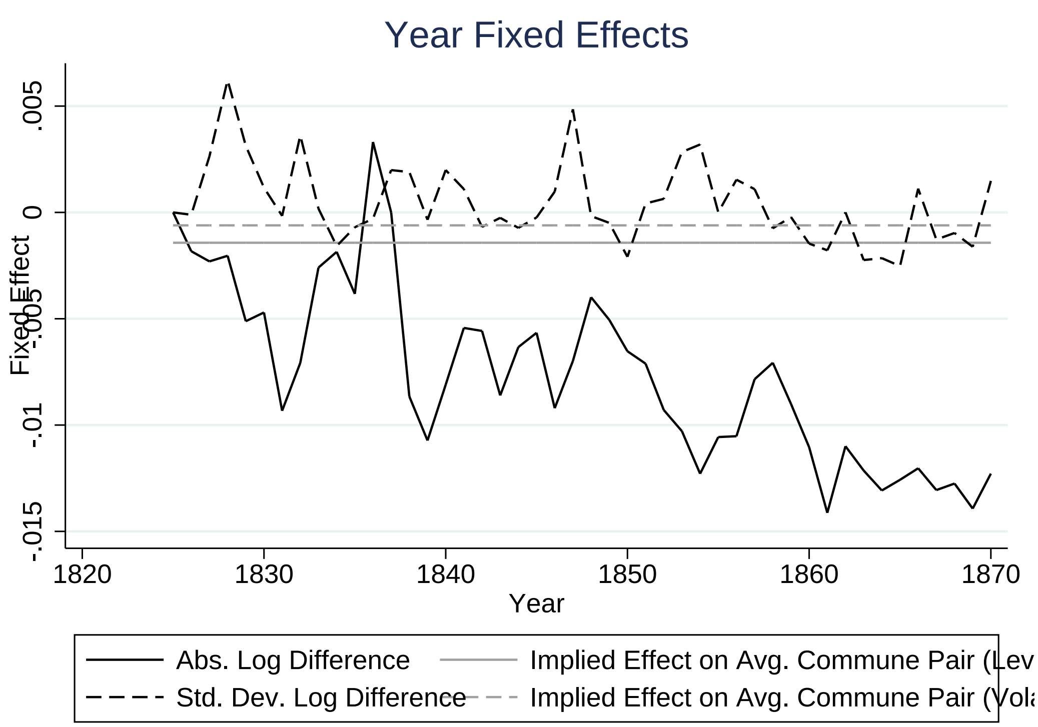
Another way to assess the importance of rails on price dispersion during this period is to compare the magnitude of the effects of rail connections to overall falls in price dispersion during the sample period. In Figure 5, I plot the year fixed-effects from the baseline regression in (1), normalized to the 1825 fixed effect equal

Table 4: Implied rail effects on hypothetical *communes*: price differential volatility

	"Average" *Commune*		"Maximal" *Commune*	
Specification	Change in $P_{i,j,t}$	Attenuation of Distance Effect	Change in $P_{i,j,t}$	Attenuation of Distance Effect
Baseline	-.0006081	4.8%	-.0012883	8.6%
$d_{min} = 100$	-.0006587	5.3%	-.0012531	8.7%
$d_{min} = 150$	-.0006983	5.7%	-.0011596	8.1%
$d_{min} = 200$	-.0006931	5.6%	-.0011092	7.7%

to 0.[6] As is apparent, the magnitude of the rail connection effect is dwarfed by the change in year fixed-effects over the sample period. One interpretation of this comparison is that while rail connections may have decreased price dispersion across distant *communes*, transportation and market changes over time that affected all *communes*, regardless of rail connections, were much more important.

Figure 5: Year fixed-effects, normalized to $\gamma_{1825} = 0$



[6]The year fixed-effects are virtually identical across specifications

4 Conclusion

References

Google Developers. The google maps geocoding api, 2016. URL https://developers.google.com/maps/documentation/geocoding/.

Soren Johansen and Katarina Juselius. Maximum likelihood estimation and inference on cointegration with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52(2):169 – 210, 1990.

Francois et Maguy Palau. *Le Rail en France: Le Second Empire, Tome 1, 1852-1857*. Imprimerie STEDI, Paris, France, 1998.

Francois et Maguy Palau. *Le Rail en France: Le Second Empire, Tome 2, 1858 - 1863*. Imprimerie STEDI, Paris, France, 2001.

Francois et Maguy Palau. *Le Rail en France: Le Second Empire, Tome 3, 1864-1870*. Imprimerie STEDI, Paris, France, 2004.

Bertrand Roehner. Wheat trade and wheat prices in france, 1486-1913, 2016. URL http://doi.org/10.3886/ICPSR09777.v1.