Max Marcussen
3/22/2020
Applied Microeconometrics
Final Project

Exit, Voice, and Political Change: Evidence from Swedish Mass Migration to the United States

A variable guide is included in the appendix.

1. Paper Introduction

The paper I chose to replicate for my final project was "Exit, Voice, and Political Change: Evidence from Swedish Mass Migration to the United States," by Mounir Karadja of Uppsala University and Erik Prawitz of the Research Institute of Industrial Economics.

The paper is concerned with the wave of migration from Europe to the US in the 1800s, and how that wave impacted political outcomes in Europe. In total, 30 million Europeans left their home countries and emigrated to the United States throughout this Age of Mass Migration. Most of this migration was sustained and continuous — family members would encourage their families to migrate, word would reach across the Atlantic about opportunity in the United States, and so on. The paper makes a preliminary observation that countries with high levels of migration in the Age of Mass Migration have a higher level of union participation in the modern era.

Noticing that Sweden had one of the highest exit rates in the Age of Mass Migration, and currently has one of the highest unionization rates, the authors chose to focus their analysis there. Over the course of 60 years, nearly 25% of Sweden's population (1.3 million) emigrated to the United States. Subsequently, in the late 1800s and early 1900s, Sweden had one of the largest labor union pushes: "The dominant force in Swedish twentieth-century politics, the Social

Democratic Party, and the powerful labor union movement were founded during the period and became key actors in reforming Swedish policy and political institutions."[1]

Initially, the authors observe that "municipalities that experienced more emigration during the Age of Mass Migration exhibit significantly stronger demand for political change" immediately after the era of migration. However, municipalities with more migration experience more leftist political climates even to this day. Welfare expenditures also increased in municipalities with higher migration. The paper points toward a few mechanisms by which increased emigration can lead to different political outcomes, drawing on previous literature to substantiate these mechanisms. First, increased migration can lead to a perception of increased outside options, thus giving labor unions more bargaining power: if workers don't like the economy, they can leave. Similarly, the option to emigrate reduces the downside risk of organizing labor: if you try to unionize and fail and get fired instead, you can simply move to America. Emigration also changed the composition of workers: in regions with higher emigration, the share of workers in the emerging industrial sector increased. There's therefore a self-selecting effect of workers that stay in Sweden — industrial workers are more likely to stay and also more likely to unionize, while agricultural workers are more likely to emigrate. Finally, Swedes that emigrate to the US likely transmit American ideals back to their countrymen at home and make people in Sweden more aware of outside options.

Though there seems to be a few mechanisms by which emigration might impact political outcomes, it might be the case that the two simply move together. Regions with higher emigration might just also have higher political change, both with the same external cause. In addition, a sudden political change might cause emigration, or a sudden event might have impacts on both — it's not entirely clear which causes which. To pin down a causal effect, the authors use frost shocks as an instrument for emigration. A sudden cold spell should cause an exogenous shock to emigration and should be uncorrelated with local politics. The authors feel that they can then use a two-stage least squares design to pin down the causal relationship between emigration and political outcomes. Frost shocks, the authors assume, have no direct

[1] Karadja and Prawitz, 1865.

impact on local politics.[2] If this exclusion assumption holds true, a frost shock's only impact on political outcome will be through the emigration it produces. We can therefore use the effect of an exogenous shock to emigration to determine the effect of emigration on local politics. In this case, emigration and political change will not have the same cause — an apolitical sudden shock to emigration will result in political change only through that emigration.

2. Data

The authors' final data set contains 1.1 million emigrants from 1867 to 1920. The data is compiled by the authors from multiple sources.

Emigration data was collected from two sources. The first was a demographic data set from the State Church of Sweden. Individual parishes would collect large amounts of population data including emigration data, which would later be centralized. The second source was archived passenger lists kept by shipping companies during the Age of Mass Migration, with self-reporting of home towns. The final data set uses the maximum value between the two data sets.

Election and labor movement data is organized in a few ways. Municipal-level voting data between 1911-1921 measures electoral participation as well as distribution of votes between political parties. The data set also includes yearly (as of December 31) local organization membership from 1881-1945, as sourced from the Social Movement Archive. These organizations include labor unions and the Social Democratic Party, the memberships of these two being grouped into a single "labor movement membership" variable. Participation in the 1909 general strike is also included.

Weather data is based on historical daily data from 32 unique weather stations, with data collected between 1864 and 1867. The median distance from centers of municipalities to weather stations is 36 kilometers. The authors acknowledge that the sparseness of weather stations might pose a problem for instrumenting emigration on frost shocks, especially if the instrument is sharp

[2] This assumption can seem dubious — why wouldn't frost shocks affect local politics? Wouldn't hunger make people less happy with the government? The authors further qualify this assumption when constructing their instrument, as I will explain below.

variation in weather. The authors assume that temperature is “evenly distributed over large areas, especially in the northern hemisphere.”[3] The assumption intuitively makes sense — it’s unlikely that a city 10 miles from a weather station will have an extremely severe cold spell compared to what’s observed at the weather station.

Finally, other historical data such as harvest yields, child mortality, and population data is included. For simplicity, I did not replicate the data compilation process. The authors provided an organized data set that was ready for instant use.

3. Empirical framework

The empirical framework we’re replicating from the paper is as follows:

$$y_{mct} = \beta \text{Emigration}_{mct} + \phi_c + \mathbf{X}'_{mc}\beta_X + \eta_{mct}, \tag{1}$$

Variables:

- y_{mct}: a political outcome in municipality m. This could be labor movement membership, electoral participation and results, or strike participation.
- Emigration_{mct}: measures emigration from 1867 to year *t*. These values are log values — this is to smooth over large spikes and to create a more linear distribution.
- ϕ_c: fixed effects for the 24 counties in the data. This is to not too broadly generalize effects of emigration on political outcome — these effects might be different in various counties.
- X_{mc}: control variables for municipality characteristics. These include: land area and arable land share; log of proximity to various economic centers like trade ports, Stockholm, or railroads; latitude and longitude; crop yields for barley, wheat, oats, and dairy; and forestation. Also included in every regression is proximity to port and 1865 population.

This is the final causal equation we’re trying to identify. Of course, we cannot directly determine this to be causal, so we have to use an instrument.

[3] Citation!

4. Using frost shocks as an instrument

The crux of the paper is using frost shocks as an instrument for emigration. The idea is that the relationship between emigration and political outcomes might not be strictly causal — one county might simply have very high emigration and be very liberal, with the two not being correlated at all. And even though the authors use fixed effects for counties, individual counties likely have their own different emigration patterns and different political outcomes, which could lead to bias. This is the rationale for using frost shocks as an instrument — shocks will push people to emigrate in a way that's similar everywhere and unrelated to individual county fixed effects. The mechanism for this push is assumed to be agricultural. Therefore, only growing season frost shocks are included in the instrument. Frost shocks are calculated by tallying the number of months where the number of frost days above average number of frost days per month was higher than the standard deviation of number of frost days per month.

Frost shocks are likely to not have similar effects everywhere. A frost shock close to a port will likely result in more emigration — people have more of an option to flee the cold. The final instrument used is therefore number of frost shocks in 1864-67 interacted with proximity to nearest emigration port. The intuition is that the marginal effect of frost shocks on emigration should be larger for counties that are close to emigration ports. Direct effects of shocks and port proximity are still controlled for and included in the regression.

The main identifying assumption behind a two-stage least squares regression is that the instrument is exogenous and independent — it has no impact on the outcome variable other than through the independent variable. Here, this is translated to frost shocks not impacting local politics beyond their effect on emigration. This assumption seems dubious at first. It would be odd if frost shocks or large agricultural shortages had no impact on local politics. The authors feel, however, that they can establish exogeneity and avoid direct effects of severe economic shocks on political outcomes by using the interaction term of port proximity and frost shocks. This term should, in theory, only be correlated with emigration: the colder it gets and the closer you are to port, the more likely you are to emigrate. The authors feel they can control for direct political impacts of frost shocks by using this term and by still including raw frost shocks in their specification. To control for various other economic impacts, the authors include interaction terms between frost shocks and various "market access" terms, namely proximity to nearby towns and trade ports.

To test the identifying assumption, the authors ran a reduced-form regression[4] of labor organization on shocks. This is Table 4 in the original paper. My replication of these results is included as Table 1 below. My regression finds a slight correlation between the instrument *shockport*, or frost shocks interacted with port proximity, and *labor_capy0020*, or labor organization rate. Standard errors, as with every regression the authors run, are clustered by weather station ID.

Table 1: Establishing exclusion of instrument with final outcome variable. The dependent variable is labor organization rate, while shockport represents our primary instrument, number of frost shocks from 1864-67 interacted with proximity to emigration ports. Also included are direct effects of shocks. Specification 2 includes local controls such as agriculture yields, arable land, and proximity to economic agents like railroads and Stockholm. Specification 3 includes frost shocks interacted with proximity to trade ports and nearby towns, while specification 4 includes non growing season frost shocks and an interaction term with port proximity for these shocks.

	1: regressing labor organization on frost shocks	2: including local controls	3: controls for market access	4: includes nongrowing season frost shocks
constant	-0.0607***	0.0182	0.0185	0.0009
	(0.0092)	(0.1898)	(0.1943)	(0.1881)
shockport	0.0014***	0.0016***	0.0017***	0.0020***
	(0.0004)	(0.0003)	(0.0003)	(0.0004)
shock	0.0001	-0.0003	-0.0005	-0.0001
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
shockport10			-0.0010*	-0.0014**
			(0.0005)	(0.0006)
shocktown			-0.0003	-0.0004
			(0.0003)	(0.0003)
shockportngs				-0.0002
				(0.0004)
shockngs				-0.0013***
				(0.0004)
R-squared	0.07	0.15	0.15	0.15
No. observations	2358	2358	2358	2358

[4] Table 4 in original paper.

A few interesting results can be taken from the exclusion test in Table 1. First, the instrument *shockport* has a positive and statistically significant relationship with labor organization. Therefore, exclusion may be violated. This relationship only goes up the more controls are added, and stays just as significant. We also see that proximity to trade ports (*shockport10*) interacted with frost shocks is negatively correlated with labor organization rate. Frost shocks' positive effect on labor organization seem to be mitigated by higher economic access. Finally, and very interestingly, non-growing season frost shocks have a strong negative impact on labor organization rates. Non-growing season frost shocks might desensitize municipalities to frost shocks generally, and might therefore reduce the impact of growing season frost shocks on labor organization rates.

Do these results show that our instrument is not excluded? Not necessarily. The strong correlation seen could simply reflect a strong correlation between frost shocks and emigration, and then between emigration and political outcome. However, it's impossible to tell right now whether this is the case. The unique extension will focus on child mortality and whether it can determine violations of exclusion.

Interestingly, many of the coefficients in my replication are slightly stronger than the coefficients in the paper. I've exactly replicated the authors' code from Python to Stata, so it's still unclear to me where this increased strength comes from. The statsmodels packages I use might have some slight coding differences with Stata.

5. Identification strategy

Overall, our identification strategy is as follows:

$$\begin{aligned}\text{Emigration}_{mct} = \gamma_{SP}(\text{Shocks} \times \text{Port})_{mc} + \gamma_S \text{Shocks}_{mc} + \gamma_P \text{Port}_{mc} \\ + \theta_c + \mathbf{X}'_{mc}\gamma_X + v_{mct},\end{aligned} \tag{3}$$

$$\begin{aligned}y_{mct} = \beta \widehat{\text{Emigration}}_{mct} + \beta_S \text{Shocks}_{mc} + \beta_P \text{Port}_{mc} \\ + \phi_c + \mathbf{X}'_{mc}\beta_X + \varepsilon_{mct},\end{aligned} \tag{4}$$

The first stage of our analysis will be to demonstrate that shocks interacted with port proximity are a strong instrument for emigration. The second stage will be to demonstrate that a reconstructed version of emigration based on frost shocks is a predictor of certain political outcomes. Shocks are still included in the second-stage regression so as to control for direct effects of frost shocks on political outcomes. Proximity to port and shocks are also included in both regressions.

6. Replication of 1st-stage OLS

The next step is to make sure that our selected instrument, frost shocks interacted with emigration port proximity, is correlated with emigration. That is, we estimate Emigration_{mct} via our instrument *shockport*. Our emigration variable here is cumulative emigration from 1867 to 1920. As in Table 1, we use four specifications:

1. No controls, only *shockport* and *shock*.
2. Adding control variables for municipality characteristics.
3. Adding economic controls such as shocks interacted with proximities to towns and
4. Adding non-growing season shocks and these shocks interacted with emigration port proximity. Essentially, a non-growing season of our instrument. This serves to validate that our instrument primarily works through agricultural means.

Proximity to port and 1865 population are once again included in every regression. Standard errors are clustered by weather station ID.

The results of this regression are included in Table 2. They find a robust correlation between *shockport* and our emigration variable, *emi67t*, in all four specifications. All four specifications also boast an extremely high R^2 of over 0.76.

<u>Table 2</u>: Replication of first-stage OLS. The dependent variable is emigration from 1867 to 1920. As evident below, all p-values and R^2 values are robust. It also seems that adding more controls does not meaningfully change the strength of the instrument. We can therefore conclude that we have a strong instrument.

	1	2: including local controls	3: controls for market access	4: includes nongrowing season frost shocks
constant	-2.7204***	3.8694	4.1610	4.2195
	(0.2861)	(4.7304)	(4.6490)	(4.6549)
shockport	0.0635***	0.0645***	0.0654***	0.0647***
	(0.0157)	(0.0138)	(0.0137)	(0.0147)
shock	0.0036	0.0106*	0.0081	0.0069
	(0.0061)	(0.0062)	(0.0084)	(0.0106)
lproxemiport	-0.1068	-0.2269***	-0.2297***	-0.2345***
	(0.0774)	(0.0776)	(0.0795)	(0.0817)
lpop1865	1.1429***	1.1886***	1.1880***	1.1881***
	(0.0386)	(0.0452)	(0.0447)	(0.0470)
shockport10			-0.0128	-0.0114
			(0.0208)	(0.0203)
shocktown			0.0020	0.0021
			(0.0045)	(0.0044)
shockportngs				0.0007
				(0.0165)
shockngs				0.0043
				(0.0138)
lproxtown		0.0618***	0.0603***	0.0601***
		(0.0193)	(0.0187)	(0.0186)
R-squared	0.76	0.77	0.77	0.77
No. observations	2359	2359	2359	2359

Results of First Stage OLS

What's most interesting about these results is that they demonstrate robustness of the instrument even without controls. Specification 1 has no controls other than port proximity, frost shocks themselves, and population. Yet the instrument is still extremely strong, with a tight standard error (0.0157 on a coefficient of 0.0635) and an extremely tight p-value ($p<0.01$). Table 2 also demonstrates that *shockport*, the interaction term between shocks and emigration port proximity, is a far better instrument than shocks on their own (*shock* has a standard error of 0.0061 which is wider than its coefficient of 0.0036). When *shockport* is included, the coefficient for shocks becomes statistically insignificant.

A bizarre result is the negative correlation between *lproxemiport*, or log of proximity to nearest emigration port, and emigration. This is especially true as more controls are added. The closer a county is to an emigration port, the less emigration it is likely to experience, independent of frost shocks.[5] This is counterintuitive — after all, if emigration is more available, people should emigrate more. There are a few plausible explanations. Cities with large emigration ports are more likely to have burgeoning economies and industrial sectors, meaning that emigration is less necessary. This is supported by the authors' claim that the primary motive for emigration was agricultural — frost shocks only push people to emigrate via their impact on agriculture, so it follows that most people who leave rely on agriculture for their livelihood, which is less true in a large maritime city. Regardless of the explanation for the negative correlation, the authors deserve criticism for not including the results for port proximity in their regression tables. The same is true of the results for population, which has a strong positive correlation with emigration. This also has a reasonable agricultural explanation — if there's a large famine, it likely feels more severe in a larger municipality where far more people die. Proximity to towns is also included in my replication and not in the authors' original table, and also has a strong correlation with emigration, likely for a similar reason.

Specification 2 includes local municipality characteristic controls such as arable land, economic access and agricultural yield. When these controls are included, the negative correlation between port proximity and emigration increases. Even when controlling for

---

[5] The variable used in the paper is proximity, which is defined as "minus the log of distance."

agriculture and economic access, distance from ports increases emigration, strengthening the idea that distance from a port even independent of agricultural or economic factors increases the likelihood of leaving Sweden. Specification 3 includes shocks interacted with proximity to trade ports, as well as shocks interacted with proximity to the nearest town. Neither of these have a significant effect, suggesting that frost shocks drive emigration independently of economic access. Finally, specification 4 suggests that non-growing season frost shocks have no statistically significant impact on emigration, even when interacted with proximity to the nearest port. This provides strong evidence for the hypothesis that frost shocks drive emigration solely through agricultural means.

In summary, we can conclude that frost shocks interacted with port proximity is a strong instrument for emigration. We also find that port proximity and population have strong impacts on emigration, while non-growing season shocks and economic access interacted with shocks have no impact on emigration. Port proximity and population should therefore be included in the final specification so as to avoid omitted variable bias.

7. Replication of 2nd-stage

Now that we've demonstrated frost shocks interacted with emigration port proximity to be a strong instrument for emigration, we try to determine whether we can use this instrument to determine the effect of emigration on political outcomes. The goal is to determine whether emigration leads to different political outcomes via a frost shock's exogenous impact on emigration. This shock, if our instrument is excluded, is independent from any factors that would move both emigration and political change. With a frost shock, the only impact on political change should be through the emigration it produces.

Once again, our final estimation equation is

$$
\begin{aligned}
\text{Emigration}_{mct} = &\gamma_{SP}(\text{Shocks} \times \text{Port})_{mc} + \gamma_S \text{Shocks}_{mc} + \gamma_P \text{Port}_{mc} \\
&+\theta_c + \mathbf{X}'_{mc}\gamma_X + v_{mct},
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
y_{mct} = &\beta \widehat{\text{Emigration}}_{mct} + \beta_S \text{Shocks}_{mc} + \beta_P \text{Port}_{mc} \\
&+\phi_c + \mathbf{X}'_{mc}\beta_X + \varepsilon_{mct},
\end{aligned} \tag{4}
$$

Estimating (3) was what we did in our first-stage OLS. Now that we have demonstrated frost shocks interacted with ports to be a strong instrument for emigration, we can use a reconstructed version of emigration based on frost shocks to determine the causal effect of emigration on political outcome.

I will replicate two tables from the authors' analysis. Table 5 demonstrates that the IV regression is more effective than standard OLS in predicting the effect of emigration on political outcome. In this section of the analysis, the authors analyze emigration until 1900, rather than until 1920 as they did for demonstrating robustness of their instrument. They then use labor organization rate between 1900 and 1920 as their dependent variable. Table 7 then uses the IV framework to estimate the impact of emigration on expenditures per capita in 1918 and 1919, as well as degree of adoption of voluntary democracy in 1919 and 1938, measuring emigration up until the year before.

Coding challenges

Because it's the language I'm most familiar with, I chose to replicate the authors' analysis in Python. Python does not have as strong support for two-stage least squares as Stata does. The statsmodels package I used didn't allow for clustered standard errors with instrumental variables. This explains my replication's difference in results from the paper's results.

Interesting results (controls, etc.) and summary

My replication of Table 5, or demonstrating the effectiveness of an IV regression, is included as Table 3. Notwithstanding the slight differences from the paper's results, the overall picture is the same: when an instrument is used for emigration, a strong correlation between emigration and labor organization is found. Additionally, in the OLS, adding controls does nothing to improve emigration's coefficient or standard errors, but in the IV regression, adding controls makes stronger the relationship between emigration and political outcome. One interesting result is that the standard errors in the IV regression are far wider than they are in the OLS regression and the R^2 is lower. This can be attributed to frost shocks not being a perfect instrument — since frost shocks don't perfectly predict emigration, variance from the reconstructed emigration is likely higher. Additionally, when frost shocks are used as an instrument for emigration, some of the confounding relationships between emigration and political outcome that might improve correlation are ironed out. This is good for our analysis — we now only get the direct effect of emigration on politics, rather than the effect of politics on emigration or the effects of a third variable on both.

Table 3: Replication of table 5 in original paper. Demonstrates the superior predictive power of our instrumental variable regression with frost shocks interacted with port proximity. The dependent variable is labor organization rate from 1900-1920.

	OLS 1: with county fixed effects only	OLS 2: with controls	OLS 3: with shocks interacted with market access	IV 1: with county fixed effects only	IV 2: with controls	IV 3: with shocks interacted with market access
constant	-0.0364***	0.0672	0.0683	-0.2337	-0.3459	-0.0727
	(0.0087)	(0.1399)	(0.1409)	(61908.6843)	(44702.7689)	(159217.5233)
emi67t	0.0096***	0.0094***	0.0094***	0.0213***	0.0248***	0.0266***
	(0.0012)	(0.0011)	(0.0011)	(0.0082)	(0.0084)	(0.0085)
R-squared	0.09	0.17	0.17	0.07	0.11	0.09
No. observations	2358	2358	2358	2358	2358	2358

My replication of Table 7 is included below as Table 4. As the authors did, I controlled for current local population at the end of emigration whenever possible. Population data were unavailable for 1917. County fixed effects, municipality characteristic controls, and market access interacted with shocks were included in every regression. My table almost exactly replicates the authors' regression results. I included in my table a few variables that were significant but the authors excluded. Proximity to nearest town, proximity to Stockholm, and barley yields were all significant predictors of expenditures per capita in both 1918 and 1919. This doesn't require much analysis — counties that have stronger economies spend more on their citizens. Overall, table 4 once again demonstrates the causal effect of emigration on political outcome — emigration strongly impacts expenditures per capita and the adoption of representative democracy.

Table 4: Replication of table 7 in paper. Dependent variables for the first two columns are expenditures per capita, and for the second two are an indicator variable for whether a municipality adopted representative democracy. All controls are hidden from the table save for ones with significant p values (p<0.01): proximity to nearest town, proximity to Stockholm, and barley yields.

	Expenditures per capita in 1918	Expenditures per capita in 1919	Representative democracy in 1919	Representative democracy in 1938
constant	-3.2076	411.9157	28.5442	-4.4989
	(1646091.9244)	(2472750.7034)	(181209.0232)	(430285.4627)
emi67t	0.9964**	1.1163**	0.0470	0.1510*
	(0.5042)	(0.4776)	(0.0350)	(0.0793)
lproxtown	0.3987***	0.2876***	-0.0051	0.0180
	(0.0957)	(0.0884)	(0.0065)	(0.0142)
lproxsthlm	0.9115***	0.8559***	0.0126	0.0114
	(0.3111)	(0.2867)	(0.0210)	(0.0494)
barley_hi	0.7410***	0.6298***	0.0133	0.0640**
	(0.2008)	(0.1846)	(0.0135)	(0.0297)
R-squared	0.09	0.09	0.06	0.21
No. observations	2219	2203	2203	2216

8. Unique extension introduction

My reduced regression in table 1 finds a slight correlation between the instrument *shockport*, or frost shocks interacted with port proximity, and *labor_capy0020*, or labor organization rate. While this doesn't definitively mean that the instrument fails the exclusion restriction, there could be causal effects of frost shocks interacted with port proximity beyond those through emigration.

The data set includes child mortality. Child mortality seems to be a good indicator of overall quality of life — it can indicate healthcare quality, education quality, and overall prosperity. Additionally, it seems susceptible to agricultural frost shocks — if a sudden cold spell decimates crop yields, mortality will likely go up. Proximity to emigration ports is also similar to proximity to urban areas and proximity to economic centers. Child mortality, given its indications of quality of life and its seeming susceptibility to frost, is likely the best candidate for a variable that's affected by our instrument in a way that's not through emigration.

9. Unique extension exploration + preliminary results

We replicate our first stage OLS, this time using child mortality in place of emigration, and still using frost shocks interacted with port proximity as our instrument or variable or instrument. This table is included below as Table 5. The same controls are used, with two significant controls, proximity to Stockholm and proximity to the nearest town, included in the table. All results are taken from specification 4, with the most controls. We find a strong negative correlation (-1.9346) between child mortality and our instrument. While emigration grows more positive with our instrument, child mortality grows more negative in a statistically significant ($p<0.01$) way that grows more significant with more controls. We also find a strong negative correlation between frost shocks and child mortality (-1.6054, $p<0.01$). More interesting, however, is how strongly correlated child mortality is with port proximity (8.0925, $p<0.01$), proximity to Stockholm (15.9785, $p<0.01$), and proximity to the nearest town (4.1410, $p<0.01$).

Table 5: First-stage OLS for unique extension. The dependent variable is child mortality rate per county. For our variable of interest, *shockport*, we find a robust negative correlation between frost shocks interacted with port proximity and child mortality. This correlation increases with more controls. We also find that proximity to ports sharply increases child mortality, as does proximity to Stockholm and proximity to the nearest town.

	1: regressing labor organization on frost shocks	2: including local controls	3: controls for market access	4: includes nongrowing season frost shocks
constant	245.6428***	-72.3194	-65.6919	-29.2498
	(32.0321)	(184.9949)	(195.9546)	(194.9129)
shockport	-1.4770	-1.7036**	-1.7004***	-1.9346***
	(1.0373)	(0.6829)	(0.5984)	(0.6390)
shock	-0.2425	-0.8386**	-1.1619***	-1.6054***
	(0.5193)	(0.3570)	(0.4242)	(0.5788)
lproxemiport	10.9997***	9.8829***	9.6566***	8.0925***
	(3.3010)	(2.2670)	(2.3130)	(2.3145)
shockport10			-1.2547	-0.7241
			(0.8486)	(0.7492)
shocktown			-0.0438	-0.0012
			(0.3214)	(0.3127)
shockportngs				0.0843
				(0.8017)
shockngs				1.5037**
				(0.6267)
lproxtown		4.3746***	4.2050***	4.1410***
		(1.2142)	(1.1937)	(1.1935)
lproxsthlm		19.3450***	18.4679***	15.9785***
		(4.5503)	(4.4582)	(4.3610)
barley_hi		-2.0853	-1.2344	-0.1334
		(3.2153)	(3.2734)	(3.4022)
R-squared	0.16	0.18	0.18	0.18
No. observations	1779	1779	1779	1779

These large coefficients lead to the conclusion that urban areas have higher child mortality. Unfortunately, the data set does not contain crime statistics. But proximity to Stockholm is the strongest predictor of child mortality. If urban areas have higher child mortality, rural areas will have lower child mortality. Additionally, rural areas are more susceptible to growing season frost shocks than urban areas (after all, they’re the regions that grow food), which means that, given our two-stage least squares results, they’re more susceptible to emigration. If frost shocks interacted with port proximity decreases mortality, rural regions are the most susceptible to emigration, and rural regions have the lowest mortality, could *shockport*'s negative impact on child mortality be through emigration? Frost shocks might simply encourage emigration, which reduces child mortality.

To test the emigration hypothesis, we first run an OLS regressing child mortality on emigration, using 1920 emigration data as the authors do in their first-stage OLS. These results are included in the appendix as Table 1a. There is a slightly positive correlation between emigration and child mortality (2.7755, standard error of 2.2265).

10. Further testing emigration-child mortality hypothesis

To test the emigration hypothesis further, we re-run our second stage OLS, this time using child mortality as our outcome variable. We use our original IV construction, using *shockport* as an instrument for emigration to test our emigration mortality hypothesis. The only change is switching the final outcome variable from emigration to child mortality. Here we find the strongest support for the hypothesis that frost shocks impact child mortality largely through emigration. The results are included in Table 6. In our OLS regressions, we find little to no correlation between child mortality and emigration (in regression with most controls, 1.1280 with a standard error of 1.9605). This mirrors our results from Table 1a. In our IV regressions, however, we find a strong negative correlation between instrumented emigration and child mortality (in regression with most controls, -22.5307 with standard error of 13.3069 and $p<0.1$). Raw frost shocks had no meaningful impact. We can conclude from this regression and from Table 5 that frost shocks primarily negatively impact child mortality through emigration.

In conclusion, child mortality, our best candidate for an economic indicator impacted by frost shocks through non-emigration means, is, in reality, mostly impacted by frost shocks through emigration. Child mortality is negatively correlated with frost shocks interacted with port proximity, but when we use this as an instrument, the correlation grows by ten-fold (-1.9346 in reduced-form regression vs. -22.5307 in two-stage least squares). This strengthens the authors' argument that their instrument only impacts economic and political outcomes through emigration.

Table 6: Replication of second-stage OLS, this time with child mortality as the outcome variable. The dependent variable is child mortality.

	OLS 1: with county fixed effects only	OLS 2: with controls	OLS 3: with shocks interacted with market access	IV 1: with county fixed effects only	IV 2: with controls	IV 3: with shocks interacted with market access
constant	244.7048***	142.8892	161.4043	484.3796	471.2652	-17.8710
	(14.4339)	(224.9800)	(226.2876)	(240396104.7920)	(nan)	(110877736.4842)
emi67t	1.4107	1.1469	1.1280	-20.3481	-22.5912*	-22.5307*
	(1.8897)	(1.9601)	(1.9605)	(13.6105)	(13.2816)	(13.3069)
R-squared	0.15	0.18	0.18	0.09	0.11	0.11
No. observations	1779	1779	1779	1779	1779	1779

11. Conclusion and overall comments

The authors have produced a robust paper with a strong instrument that holds up under further testing. In my replication, I demonstrated the replicability of the authors' results and largely confirmed the causal effect the authors found. I also added variables to tables that had interesting correlations and that the authors omitted. Overall, my replication demonstrates a strong causal effect of emigration on political outcome. In my unique extension, I provided more support for the authors' identifying assumption by analyzing a variable that I thought might be impacted by frost shocks in a way that wasn't through emigration. I ended up demonstrating that emigration explains most of the change in child mortality due to frost shocks. All of the empirical evidence suggests that emigration strongly impacts political outcomes, with a causal effect pinned down by our two stage least squares approach.

The greatest difficulty in replicating the authors' paper was in coding. The authors provided extensive Stata code that I translated into Python. Unfortunately Python does not have strong support for two-stage least squares regressions.

Further analysis could include further verifying the authors' instrument by trying to pin down other effects of frost shocks. Another interesting project would be to see whether areas in America that experienced more immigration from Europe also experienced different political outcomes. Since most of the emigration was agricultural, it would also be interesting to study if most emigrants from Sweden ended up in rural areas of the United States.

Appendix

Table 1a: child mortality regressed on emigration, with standard controls. The dependent variable is child mortality.

	1: regressing labor organization on frost shocks	2: including local controls	3: controls for market access	4: includes nongrowing season frost shocks
constant	254.3002***	-94.7419	-91.2365	-67.9428
	(36.0548)	(190.8984)	(199.3168)	(200.9419)
emi67t	3.2553	2.9277	2.8670	2.7755
	(2.1596)	(2.1055)	(2.1516)	(2.2265)
shock	0.1950	-0.3709	-0.6818	-0.9348
	(0.7135)	(0.4735)	(0.5007)	(0.5882)
lproxemiport	14.2494***	13.9565***	13.7395***	13.1471***
	(2.9818)	(1.9067)	(1.9034)	(2.0229)
shockport10			-1.1910	-0.7714
			(0.8813)	(0.7581)
shocktown			-0.1045	-0.0617
			(0.3382)	(0.3246)
shockportngs				-0.3946
				(0.9418)
shockngs				1.0054
				(0.6596)
lproxtown		4.1298***	3.9626***	3.9667***
		(1.2396)	(1.2164)	(1.2176)
lproxsthlm		18.8769***	18.0392***	15.4263***
		(4.6478)	(4.9629)	(5.2186)
R-squared	0.16	0.18	0.18	0.18
No. observations	1779	1779	1779	1779

List of variables:

emi67t: emigration from 1867 to 1920.

shock: frost shocks between 1864 and 1867.

shockport: proximity to nearest emigration port.

lproxemiport: proximity to local emigration port.

lpop1865: population as of 1865.

shockport10: proximity to local trade port.

shocktown: proximity to nearest town.

shockngs: non-growth season shocks.

shockportngs: shocks in non-growing season interacted with proximity to emigration ports.

lproxtown: proximity to town.

lproxsthlm: proximity to Stockholm.

barley_hi: barley yields.

Citations

Karadja, Mounir, and Erik Prawitz. “Exit, Voice and Political Change: Evidence from Swedish Mass Migration to the United States.” *Journal of Political Economy*, June 2019. https://www-journals-uchicago-edu.proxy.uchicago.edu/doi/full/10.1086/701682.