

Karadja, Mounir; Prawitz, Erik

Working Paper

A response to Pettersson-Lidbom's "Exit, voice and political change: Evidence from Swedish mass migration to the United States - a comment"

Working Paper, No. 2020:5

Provided in Cooperation with:

Department of Economics, Uppsala University

Suggested Citation: Karadja, Mounir; Prawitz, Erik (2020) : A response to Pettersson-Lidbom's "Exit, voice and political change: Evidence from Swedish mass migration to the United States - a comment", Working Paper, No. 2020:5, Uppsala University, Department of Economics, Uppsala,
<https://nbn-resolving.de/urn:nbn:se:uu:diva-417028>

This Version is available at:

<https://hdl.handle.net/10419/249374>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



UPPSALA
UNIVERSITET

Department of Economics
Working Paper 2020:5

*A response to Pettersson-Lidbom's "Exit, Voice
and Political Change: Evidence from Swedish
Mass Migration to the United States – a
Comment"*

Mounir Karadja
Erik Prawitz

Department of Economics
Uppsala University
Box 513
751 20 Uppsala
Sweden

Working Paper 2020:5
August 2020
ISSN 1653-6975

*A response to Pettersson-Lidbom's "Exit, Voice and
Political Change: Evidence from Swedish Mass Migration
to the United States – a Comment"*

*Mounir Karadja
Erik Prawitz*

A response to Pettersson-Lidbom’s “Exit, Voice and Political Change: Evidence from Swedish Mass Migration to the United States – a Comment” *

Mounir Karadja[†]

Erik Prawitz[‡]

August 7, 2020

Abstract

In a comment to Karadja & Prawitz (2019), henceforth KP, Per Pettersson-Lidbom (2020), henceforth P-L, argues that the main results in KP are severely biased. He argues that KP’s results are biased due to non-classical measurement error in emigration and due to confounders related to the instrument.

In this response, we show that P-L’s reasoning regarding measurement error bias contradicts the results from his proposed test. More generally, P-L’s results cannot exclude alternative and arguably more likely explanations. We present two straightforward tests that both indicate that measurement error does not bias KP’s results.

Second, we argue that KP controls for confounders in a standard way given the identification strategy. Including fixed effects at the level of the exogenous cross-sectional variation, as P-L does, severely limits the available identifying variation and decreases precision. Nevertheless, we document that KP’s results are robust to non-linear frost shock controls, including fixed effects for groups of similar frost shocks. In addition, we show that our results are robust to altering regional fixed effects or dropping them altogether, in contrast to what is suggested by P-L.

*We thank numerous colleagues for helpful comments and suggestions. We have also benefitted from the insightful comments by the editor Magne Mogstad and three anonymous referees on Pettersson-Lidbom (2020), which was rejected for publication in the *Journal of Political Economy*.

[†]Department of Economics, Uppsala University.

[‡]Research Institute of Industrial Economics.

1 Introduction

In a recent comment, Pettersson-Lidbom (2020) (henceforth P-L) criticizes different aspects of our paper “Exit, Voice, and Political Change: Evidence from Swedish Mass Migration to the United States”, published in the *Journal of Political Economy*, vol. 127, no. 4 (henceforth KP). Using an instrumental variable design, our paper shows that Swedish emigration during the Age of Mass Migration bolstered the nascent labor movement at home. Due to considerable chain migration, we show that our proposed instrument, the interaction of local frost shocks 1864–67 and proximity to an emigration port, predicts emigration during several subsequent decades.

P-L presents two main points of critique. First, he argues that emigration is underreported in the data used by KP, that the ensuing measurement error is non-classical, and that the results in KP are biased as a consequence. P-L also proposes a method to correct for the argued bias using own data on outmigration. Second, he argues that KP does not account for confounders related to the weather-based instrument, such as non-linearities in the effects of frost shocks and other unobserved heterogeneity at the weather-station level. In particular, P-L suggests to include fixed effects for all unique values of frost shocks to control for non-linearities and to include weather-station fixed effects to control for related confounders. Ultimately, P-L argues that his critique reveals that there is no causal relationship between emigration and the labor movement.

Regarding the first point of critique, we start by noting that it is likely that emigration was underreported to some extent, as already discussed in KP. However, we find that P-L provides a selective reading of the literature on the Swedish emigration records and exaggerates concerns on the underreporting of emigrants. More importantly, we explain why P-L’s proposed test is ill-suited to test for the presence of measurement error bias. Instead, we present two straightforward tests documenting that measurement error in emigration is unlikely to bias our results. In our preferred test, we document that our results are robust to only considering emigration during 1885–1893, a period known to have little unrecorded emigration and hence low measurement error (see, e.g., Sundbärg, 1913; Bohlin & Eurenus, 2010). We also note that P-L’s own reasoning regarding the bias due to measurement error contradicts the results of his proposed test. In particular, P-L argues that underreporting of emigrants in KP leads to an upward bias in *both* the first-stage and the second-stage estimates. However, since the second-stage IV estimate is the ratio of the reduced form to the first-stage coefficients, it is generally not possible to overestimate both relationships.

Regarding the second point of critique, we note that P-L misunderstands our stated regression model, and motivate why P-L’s suggested specifications do not impair the results in KP. We argue that P-L’s claim that “allowing that the effect of the weather shock on the outcome of interest may be non-linear shows that there is no first-stage effect” is incorrect.

In fact, KP already implements several specifications where frost shocks are allowed to have non-linear effects on outcomes. In this response, we complement those results with specifications that include fixed effects for municipalities that had very similar exposure to frost shocks.

We note that P-L's related suggestion to include weather-station fixed effects is similar but even more demanding than including frost shock fixed effects. Given the cross-sectional variation at hand, P-L's suggested specification severely limits the available identifying variation, which is reflected in a decreased precision. Although estimates also decrease when we add weather-station fixed effects to our baseline specification, we cannot reject that point estimates are the same as in KP's baseline specifications. In general, while P-L points at some limits to our identifying variation, we argue that KP controls for confounders in a standard way given the identification strategy and that we are nonetheless able to address concerns about non-linear confounders in a range of auxiliary specifications.

Finally, we show that the baseline results in KP are robust to controlling for different regional fixed effects. Moreover, in contrast to what is suggested by P-L, results are also robust to dropping regional fixed effects altogether, as long as we use our preferred specification where we compare observations that are similar on observables. However, this is masked by the fact that P-L only displays the result from using a model stripped from all controls except frost shocks, proximity to port, and baseline population.

The remainder of this response discusses these issues in detail. It is organized by following the points of critique presented in P-L.

2 Critique regarding underreporting of emigrants

We begin this section with a brief background on how KP measures emigration, before commenting on the extent of underreporting of emigrants. We then turn to its potential consequences for bias and explain why P-L's attempt at combining data on internal and external migration is flawed as a method to study bias due to measurement error. Finally, we perform two straightforward tests that both show that measurement error in emigration is unlikely to bias the results in KP.

2.1 Background

In KP, we measure emigration from Sweden using two sources: the emigrant lists from parish church books and the passenger lists kept by shipping companies. The former originates from the records of the State Church in Sweden, which was historically tasked with tracking demographic statistics in its parishes. The digitized emigrant data from this source used in

KP cover the years until 1895.¹ The latter originates from shipping companies with ships to foreign destinations, which were required by law to compile lists of all their passengers. The passenger list data used in KP is obtained from the *2006 Göteborgsemigranten Emigranten Populär Database* and cover the years from 1869 up until 1920.

These sources provide us with two independently collected yearly measures of emigration, with different strengths and weaknesses. While the church book data were collected at the place of origin by the parish priests, the main concern is that they do not generally include migrants without a change-of-address certificate. As we discuss below, this is primarily an issue for the earlier year of our sample period. In fact, the Emigration Ordinance of 1884 made it mandatory for emigrant agents to present a change-of-address certificate to the authorities (see, e.g., Bohlin & Eurenus (2010)). On the other hand, since shipping companies were bound by law to create passenger manifests, the passenger list data may include migrants who were not registered in the church books. However, a drawback of the passenger list data is that passengers often reported their place of origin imprecisely, making it more challenging to link them to their municipalities of origin.

From these two series, we construct one single yearly measure of emigration by taking the maximum of either the church book or the passenger list data each year.² As we note in KP, the reason for this is twofold. First, we lack emigration from the church book data towards the end of our period (after 1895), and we lack data from the passenger lists before 1869. Second, the different concerns of underreporting is likely mitigated by combining both sources and using the maximum of either source.

The final data set encompass 1.1 million migrants in total for the years 1867–1920. In our main cross-sectional regressions at the municipality level, we take the log of the number of emigrants over the entire period 1867–1920 and control for the log of population at baseline.

2.2 Underreporting of emigration

P-L remarks that it is well known that the Swedish emigration was severely underreported and claims that the emigration variable in KP accounts for “at most 73% of all emigrants during 1860–1920”. The latter claim is obtained by taking the difference between P-L’s own estimate of the true number of emigrants during the relevant period, which is 1.5 million, and the 1.1 million emigrants in KP’s data set. While P-L does not provide details on how he estimates the 1.5 million figure, he discusses three reasons for why KP’s measure of emigration is underreported: (*i*) missing migrants for unspecified reasons as compared to the

¹The data were digitized by the *Swedish Migration Center* in Karlstad.

²In general, the correlation between the two data sets is very high, as detailed in Online Appendix section A.1. For years when we have both sources, the parish data report strictly more emigrants 55 percent of the time, while the passenger data indicate strictly greater numbers 25 percent of the time. In 23 percent of cases, the data sets are equal.

official statistics (i.e. church books), (ii) parish records (i.e. church books) only registering individuals with a change-of-address certificate, and (iii) severe underreporting of emigration to destinations within Europe (most notably Denmark and Germany) in both the emigrant lists of the church books and the passenger lists.

For the first source of underreporting, P-L remarks that there are 1.3 million emigrants in the church books according to official statistics, and assumes that we are missing at least 0.2 million from these as there are 1.1 million emigrants in KP's data. Although the digitized emigrant lists of the church books used in KP may miss some migrants from the original source, the 1.3 million claim is uncertain. In fact, the figures attributed to the church books range between 1.1 and 1.3 million emigrants during the period (see, e.g., Sundbärg, 1913; Tedebrand, 1976; Ljungberg, 1997; Bohlin & Eurenus, 2010). For the other two sources of underreporting, P-L estimates that there are an additional 0.2 million emigrants missing "due to various sources of errors discussed by Johansson (1976) and Eriksson (1969), among others". No further details are provided, making it difficult for us to evaluate this precise estimate, but we discuss below the potential underreporting in KP related to concerns (ii) and (iii) mentioned above.

To begin with, however, it is worth noting that the extent of underreporting suggested by P-L is above the estimates currently in the literature. We find that P-L provides a selective reading of both the studies cited by him, as we return to below, and the literature on the Swedish emigration records more generally. While referring to a handful of studies, and criticizing that these are not cited by KP, P-L does not mention the fact that the cited work consists of small case studies, covering a few parishes, towns or counties, and that they mainly focus on a few years in the beginning of the mass migration period. For example, Eriksson (1969) examines only two southern counties for the single year of 1874. Odén (1964) studies urban migration during 1840–1872, a time period which only spans a small fraction of the emigration studied in KP.³ Similarly, Johansson (1976) studies 21 towns during the period 1860–1870. Ahlqvist (1976) studies emigration to Denmark from three Swedish parishes. Vernersson Wiberg (2016) studies emigration to Germany and Denmark from the southern county of Blekinge. In contrast, other historical case studies finding the emigrant lists in the church books to be accurate are not mentioned in P-L (see Tedebrand, 1972; Norman, 1974; Kronborg & Nilsson, 1975).

Among the references in P-L, it is only Sundbärg (1913) that makes an attempt at looking at larger volumes of data in the official emigration report from 1913, where he studied the level of underreporting by examining net migration flows. He concludes that, while the unreported migration was about 10-20% in the period before 1884, it was substantially less

³Odén (1971) is rather an overview of the theoretical literature on emigration and does not provide any new evidence on the extent of measurement error in the Swedish emigration records.

afterwards, with as little as 1% in 1885-1893 (a period of substantial migration). In more recent work on the determinants of emigration, Bohlin & Eurenus (2010) remark that although the official statistics underestimated emigration before the mid 1880s, “the problems with the official statistics were not as large as for example Gustav Sundbärg believed”. We also note that Tedebrand (1976, pp. 84–94), not cited in P-L, argues that the emigration statistics are, after all, to be considered exceptionally good in an international perspective.

While the suggested extent of underreporting is questionable, P-L is most likely correct in remarking that emigration was undercounted. However, this issue is in fact already acknowledged in KP. Although we admittedly should have provided explicit references to, in particular, the work of Sundbärg (1913), which we cite several times elsewhere in KP, our reading of the literature is essentially the same today as it was at the time of writing KP. In particular, as we will discuss extensively in the next section, we do not find support for the view that underreporting affects our estimates.

Moreover, while KP never claims to recover all non-registered emigrants in the church books, P-L does not recognize that underreporting is mitigated by calculating emigration as the maximum of either the church book or the passenger list data each year. For instance, a migrant without a change-of-address certificate not recorded in the church books may be found in the passenger lists. While this procedure does not work when the same migrant is missing in both sources and is less effective if both sources are underreported in the same year, we find that P-L exaggerates the extent of such concerns.

First, P-L wrongly claims that the passenger list data did not record emigrants to other destinations than the US. In fact, all passenger ships were required by law to register their passenger lists (Clemensson, 1996). And it is clear from tabulating the data that other destinations than the US are frequent, albeit few relative to the US totals, as to be expected.⁴

Second, although it is reasonable to believe that emigrants to some parts of Europe did not utilize the passenger ships, in particular migrants to neighboring Denmark, it is primarily in the early years that these migrants were misreported in the church books, as noted above. In particular, the deficiencies refer to the 1860s, as noted by, e.g., Sundbärg (1913) as well as Ahlqvist (1976). Thus, the emigration period studied in KP (1867–1920) is less affected. Moreover, the non-registered migration to Europe was to a large part seasonal in nature (see, e.g., Johansson, 1976). It is therefore not evident how it should be compared to the transatlantic emigration, which is the focus in KP. In fact, even the registered emigration to Europe may have had different economic impacts than emigration to the United States, an issue we will return to when discussing the potential bias in the second-stage relationship.

Third, P-L remarks that it is problematic that KP lack church book data after 1895. However, the bulk of emigrants left before 1895. In Table 1 we display our main results in

⁴Available at www.ancestry.se/search/collections/1189.

KP when varying how we count emigration using different sources and periods. Panel A displays first-stage estimates and Panel B displays the IV-estimated effect of emigration on labor movement. For reference, we display our baseline measure of emigration in column 1. While we postpone the discussion on bias due to measurement error to the next section, we note here that results are stable when considering only the period up until 1895.

In practice, the method of taking the maximum of each source have little consequence for the estimates when compared to using only the church book data. This is seen when comparing our baseline results to columns 3 and 4, where we document the results when using only the church book data or the passenger list data for the available years. In the next section, we turn to the consequences of measurement error for bias in more detail.

2.3 Potential bias due to measurement error

More importantly for the results in KP, P-L argues that KP's main estimates are biased due to underreporting of emigration. We evaluate such concerns in detail here.

To fix ideas, suppose that we want to estimate the effect of emigration on an outcome Y_i . However, emigration is mis-measured because of underreporting, such that observed emigration is given by $X_i = \theta_i X_i^*$, where $0 < \theta_i < 1$ and X_i^* is the true value of emigration. Moreover, suppose that we use an instrumental variable strategy and regress the underreported X_i on an instrument Z_i as our first-stage regression. If we, for simplicity, start with the case when θ_i is constant for all i , it is straight-forward to derive the bias. In particular, the first-stage coefficient of the instrument will be attenuated proportionally to θ_i . What about the second-stage IV coefficient? The potential bias is perhaps most clearly seen when expressing the IV coefficient as the ratio between the coefficient from the reduced form regression of Y_i on Z_i and the coefficient from the first-stage regression of X_i on Z_i :⁵

$$p \lim \beta_{IV} = \frac{Cov(Y, Z)}{Cov(X, Z)} = \frac{Cov(\beta X^* + u, Z)}{Cov(\theta X^*, Z)} = \beta \frac{1}{\theta}, \quad (1)$$

where u_i is the error term in the second-stage regression and we assume $Cov(u, Z) = 0$.⁶ While the reduced form regression of Y_i on Z_i is not biased due to measurement error in X_i , for natural reasons, the IV coefficient overestimates β with a factor $1/\theta$, due to the attenuated bias in the first-stage regression. Thus, the bias of the first-stage and the second-stage coefficients will go in the opposite directions.

Next, we investigate the case of using the natural logarithm of emigration, as in KP, rather than emigration in levels. The measurement error then becomes additive:

⁵We simplify the exposure by disregarding other regressors here.

⁶If $Cov(u, Z) \neq 0$, the exclusion restrictions would be violated. We discuss related issues in the next section.

$$\tilde{X}_i = \tilde{X}_i^* + e_i,$$

where $\tilde{X}_i = \log X_i$, $\tilde{X}_i^* = \log X_i^*$, and $e_i = \log \theta_i$. As a consequence, and in contrast to equation (1), the estimate of the first-stage coefficient is no longer necessarily biased. This is illustrated in column 5 of Table 1, where we inflate the number of emigrants by a factor 1.3699 (assuming $\theta = 0.73$), to reflect P-L’s concern that the data in KP only contain 73% of all emigrants.⁷ Instead, in order to have non-classical measurement error, we must assume that the measurement error is correlated with the instrument. Similarly, the estimate of the IV coefficient is no longer necessarily biased due to underreporting, since any potential bias still hinges on the bias in the first-stage regression:

$$p \lim \beta_{IV} = \beta \frac{Cov(\tilde{X}^*, Z)}{Cov(\tilde{X}^*, Z) + Cov(e, Z)}. \quad (2)$$

If $Cov(e, Z)$ is zero, β_{IV} is unbiased, but if it is non-zero, β_{IV} is biased. Moreover, the sign of the bias will depend on the sign of $Cov(e, Z)$.

P-L claims that $Cov(e, Z)$ is most likely negative. Since e_i measures the log share of reported emigrants, and not the number reported (or non-reported) emigrants, we note that a negative covariance must be driven by a negative correlation between the *share* reported emigration and the instrument. Assuming that \tilde{X}^* is positively related to Z_i , one could claim $Cov(e, Z) < 0$ by arguing that places with higher emigration have a systematically lower share of reported emigration. A clear argument for why this would be expected is, however, lacking in P-L.⁸ In fact, Johansson (1976), one of the main references in P-L, cannot detect any pattern between the size of “true” emigration (using a preferred source) and the share reported emigration (in the emigrant lists of the church books). Instead, he concludes that the measurement error probably depended on carelessness by the parish priests.

If we nevertheless assume $Cov(e, Z) < 0$, we can infer the sign of bias from equation (2) above. Since $Cov(\tilde{X}^*, Z) > 0$, the first-stage coefficient would be attenuated and biased *downwards*, as seen by the expression in the denominator.⁹ As a consequence, the second-stage would overestimate the effect and, in the case of $\beta > 0$, would be biased *upwards*.

⁷Note that the coefficients are not exactly the same. This is due to the fact that we add a constant equal to one before taking the natural logarithm of the number of emigrants (since there are a couple of municipalities with zero emigration 1867–1920).

⁸P-L argues that: “In the case of underreporting of emigrants, this correlation will most likely be negative since there cannot be any underreporting if there is no emigration (i.e., emigration cannot be negative!), whereas there will be underreporting if there is emigration, and the greater the size of emigration is, the higher the underreporting.” We note that only 3 out of 2,359 municipalities in KP’s data have zero emigration 1867–1900. Moreover, we cannot find that Stephens & Unayama (2019) make a similar point, in contrast to what is claimed in P-L.

⁹In theory, if $|Cov(\tilde{X}^*, Z)| < |Cov(e, Z)|$, the first-stage coefficient could even switch sign.

Indeed, P-L argues that the IV coefficient will be biased upwards due to $Cov(e, Z) < 0$. We note here, for future reference, that the bias in the first stage will go in the opposite direction of the bias in the second stage (similarly to the case in equation (1)).

Next, after arguing for $Cov(e, Z) < 0$ and an upward bias in β_{IV} , P-L proceeds with his proposed test, claiming that it solves the argued measurement error problem. More specifically, P-L adds own data on the number of total outmigrants from a parish and uses this in place of KP's emigration data. P-L argues that since he observes total registered outmigration, he observes the sum of true emigration, X^* , and internal outmigration, I . Assuming that $Cov(I, Z) = 0$ – a point we will return to below as KP shows that this is non-zero – replacing KP's emigration variable with P-L's total outmigration variable should thus produce a consistent, albeit less precise, first-stage estimate.

After regressing the total outmigration measure on the instrument, conditional on controls, P-L concludes that KP suffers from an upward bias in the first-stage relationship (“with a factor of more than 6”), in addition to the upward bias in KP's second stage. Recalling the reasoning above, he thus argues for an *upward* bias due to underreporting in emigration in *both* the first-stage and the second-stage relationships. However, since the IV estimate is the ratio of the reduced form to the first-stage coefficients, we cannot overestimate both relationships. Clearly, the reasoning in P-L is contradictory and the results from the proposed test cannot be used to conclude that measurement error is driving an upward bias in our IV-coefficients.

It is not clear what the merits of P-L's test are in the first place. If we assume that internal migration is unrelated to the instrument, combining emigration and internal migration should add noise to our first-stage estimates. Possibly a lot of noise.¹⁰ While this does not imply bias, it is far from evident that it solves any potential non-classical measurement error. Although P-L's data could be better in some regards than KP's, this cannot be taken for granted. For instance, since P-L's data originates from church books it is unclear how they can resolve the issues related to emigrants without a change-of-address certificate. Moreover, as P-L has not been willing to share how he links municipalities in his own migration data with ours, it is hard to evaluate the accuracy of these links or other measurement issues in his data.

More importantly, however, even if reliable internal migration data would be available, taking the sum of emigration and internal migration over several decades is a far-fetched exercise to test for measurement error and one that disregards the dynamic relationship

¹⁰Although it is striking that P-L reports that 1.1 million emigrants in KP's data only make up 7.6% of total outmigration on average in his data, it is unclear why this “underscores the problem of underreporting of the Swedish emigration in KP's data”. A simple back-of-the-envelope calculation gives at hand that P-L has about 14.5 million outmigrants in his data (in a population of, roughly, 4–5 million people during the period). It seems arguably more reasonable to believe that a migrant in P-L's data is something qualitatively different than a migrant in KP's data.

between internal and external migration. In particular, migration decisions are likely a function of past migration decisions in a location, since social networks abroad affect the costs of migrating. In locations with more early US emigration, it became relatively less costly to emigrate to the US, as compared to locations with low early US emigration. Thus, prospective migrants in well-connected locations were relatively *more* likely to chose the US as their destination, and correspondingly *less* likely to migrate within Sweden.¹¹ In other words, emigration and internal outmigration are likely substitutes.

In line with this reasoning, KP already documents that the instrument has opposite effects on emigration and internal migration (as measured by the number of individuals living outside of their birth municipality in a given census year). In fact, this is also noted in footnote 9 of P-L. Thus, given that KP's instrument is positively related to emigration and negatively related to internal outmigration, combining the two measures will clearly lead to a smaller point estimate as compared to the coefficient for emigration alone.

If the first-stage effect on total outmigration is weaker, or in the extreme case even close to zero, a relevant question is why we see a second-stage effect. A plausible answer is that the economic effects of emigration to the United States are different as compared to those of internal migration within Sweden. While we cannot fully exclude that internal migration may have effects on the labor movement by its own in KP, we evaluate different mechanisms behind the increase in labor movement. For instance, emigrants may transmit information or direct economic means from their destination. Since cultural aspects may have been different in the United States as compared to Sweden and since living standards were higher in the former during the 19th century, it is likely that the economic impact differed depending on destination. Similarly, the outside option of leaving for the United States for a prospective migrant may be substantially different than going to a neighboring Swedish municipality or to a Swedish town. And not only in terms of the higher expected income overseas. In fact, workers could be blacklisted for being union members (see, e.g., Westerståhl, 1945). For these reasons, prospective migrants with a stronger US outside option may have been willing to take higher risks. Ultimately, this is, however, a discussion related to the mechanisms behind the results found in KP.

2.4 Two straight-forward tests

To assess the potential bias due to measurement error, we instead perform two straight-forward tests. In particular, they are directly related to the two main issues regarding

¹¹This is also much in line with anecdotal evidence describing how locations with high emigrations saw the migration decision as a choice between America and staying at their location of origin. As noted in KP, one labor organizer in Ljusne (who had been fired for his activism) remarked that “oddly enough, there [are] only two places in the world for us, Ljusne or America”.

underreporting presented in P-L: emigrants to neighboring European locations and emigrants without a change-of-address certificate.

The first test takes its starting point in the literature that P-L mainly refers to, which focuses on non-registered emigration to Germany and Denmark in the church books. Although the instrument, the interaction of frost shocks 1864–67 and proximity to an emigration port, is non-related to latitude, as documented in Table 2 of KP, it could in theory be correlated with unreported emigration to these destinations.¹² In other words, the instrument could potentially be correlated with θ_i , as specified above. Since it is likely that emigration to Germany and Denmark, neighboring the south of Sweden, was mostly relevant for southern Sweden, this suggests a straight-forward and direct test.¹³ More specifically, we can drop southern municipalities and study to what extent our results are altered when using a subsample with supposedly less underreporting. Figure 1 provides the results from this exercise. It displays the coefficients from separate regressions where we sequentially, for each 5th percentile, have dropped municipalities that are most southern based on the latitude of their centroids (up until almost half of the sample). As seen in the figure the coefficient on emigration is stable throughout. The IV estimate of emigration on labor movement participation is similarly stable.

Why is the effect of the instrument on emigration almost unaffected by dropping southern municipalities where European migration is likely more common? At least three potential (non-exclusive) explanations emerge. One is simply that underreporting is small. Note that the studies discussing mismeasured emigration to Europe focus on the 1860s, while migration during the bulk of the Age of Mass Migration is most likely better measured. Another possibility is that unreported European emigration is unrelated to the instrument. This could be because the construction of the instrument makes it related to US emigration – the main focus in KP – and not to European migration, or because misreporting is non-related to total emigration. In fact, as discussed above, the latter would be consistent with the view in Johansson (1976).

While the test above focuses on emigration to Europe, most commonly to Germany and Denmark, emigration may still be underreported in the entire country due to missed migrants without a change-of-address certificate. While our use of the passenger lists should mitigate this type of underreporting in the church books it may not be enough. To evaluate such concerns, a more general test is suggested by the findings of Sundbärg (1913) regarding the time variation in the general underreporting of emigration. In particular, Sundbärg (1913) argues that while the unreported migration was substantial in the period before 1884, it was

¹²Note that although southern locations have fewer frost days, as compared to the north, our measure of *frost shocks* adjusts for the long-term mean and standard deviation of frost incidence. See Section IV in KP.

¹³In fact, the literature that P-L cites, such as Eriksson (1969), Ahlqvist (1976) and Vernersson Wiberg (2016), takes this aspect as a starting point when focusing on case studies of southern parishes.

as little as 1% in 1885-1893, as noted above. Thus, if we limit the period which we use for counting emigrants to 1885–1893, we can compare this estimate to the estimate obtained using the full period.¹⁴ The last columns of Table 1 provides the results from such a test. Recall that column 1 in Panel A restates the main first-stage specification used in KP, with the log of the number of emigrants 1867–1920 as the outcome variable. We can compare this estimate to the estimate in column 6, where emigration is counted during a period of supposedly higher measurement error in the church books (1867–1884), as well as to the estimate in column 7, where emigration instead is counted during a period of supposedly low measurement error in the church books (1885–1893). If anything, the estimate in column 6 is higher. In any case, it is difficult to make the case that the first-stage results are biased in a severe way due to measurement error in emigration. Columns 6–7 in Panel B display the second-stage relationship with labor movement participation as the outcome variable for the different periods of emigration. Again, coefficients are similar in magnitude across the board. Together, these tests show that potential measurement error in emigration does not severely bias the results in KP, contrary to P-L’s claims.

3 Critique regarding the correct way to control for frost shocks

In the second point of critique, P-L criticizes the identification strategy in KP and suggests to control for frost shocks in a non-linear way and for the inclusion of weather-station fixed effects. In response, we argue that P-L’s specifications severely limit the available identifying variation, which is reflected in an increase in standard errors. We also note that P-L’s claim that both the frost shock variable and its related coefficient are wrongly indexed is incorrect. Instead, we argue that KP control for confounders in a transparent and standard way given the identification strategy.

3.1 Issues with indexing

P-L begins section 2 by criticizing the indexation used in KP and argues that this makes it difficult to evaluate KP’s empirical approach. For clarity, we start by restating the first-stage and second-stage regression equations used in the main analysis of KP, as expressed on page 1885:

¹⁴Note that by choosing 1893 as the end year of the period, we use emigration data from both of our available sources (church books and passenger lists), which should also reduce underreporting.

$$Emigration_{mct} = \gamma_{SP}(Shocks \times Port)_{mc} + \gamma_S Shocks_{mc} + \gamma_P Port_{mc} + \theta_c + \mathbf{X}'_{mc} \gamma_X + v_{mct} \quad (3)$$

$$y_{mct} = \beta E\widehat{Emigration}_{mct} + \beta_S Shocks_{mc} + \beta_P Port_{mc} + \phi_c + \mathbf{X}'_{mc} \beta_X + \varepsilon_{mct}, \quad (4)$$

where m is a municipality and c is a county. While these are cross-sectional regressions, t denotes that emigration is counted from 1867 up until year t and that the outcome variable y_{mct} is measured in t .

P-L makes two remarks regarding the use of indexes in the above equations. First, he finds it noteworthy that there is a subscript S in β_S . According to P-L, this suggests that KP estimates different coefficients for each weather station, despite the fact that KP estimates only one coefficient for the variable $Shocks_{mc}$. This is a misunderstanding of KP's notation. In both of the above equations, the *gammas* in the first stage and the *betas* in the second stage are subindexed with the capital letter belonging to the name of its related variable. For example, γ_S is the first-stage coefficient belonging to $Shocks_{mc}$ and γ_P is the first-stage coefficient belonging to $Port_{mc}$. Thus, the subscript S refers to shocks and not to weather stations, as misunderstood by P-L. Even if one would misunderstand the index notation in this instance, it is arguably clear from the related text that KP does not allow the effects of $Shocks_{mc}$, $Port_{mc}$, or $Shocks \times Port_{mc}$ to vary at the weather-station level.

The second point about indexes raised by P-L is that KP uses the wrong index on the variable $Shocks_{mc}$, "since it is measured at the weather station level s and not at the county level c or municipality level m ". Again, this is a misunderstanding. On page 1881, KP notes that there is in fact some variation in shocks at the municipality level, which is why $Shock_{mc}$ is the correct index. The reason why there is some (albeit limited) variation at the municipality level is that frost shocks constructed using deviations from long-term means and standard deviations of frost incidence. Since these long-term values are collected using periods in which there are many more weather stations available, this introduces additional variation even between municipalities that share the same weather station in the 1864–67 period. Thus, while two municipalities that share the same 1864–67 weather station will always have the same incidence of *frost*, they may not have the same number of *frost shocks*, because the mean and standard deviation of frost may differ between the municipalities. Consequently, there is indeed some variation of frost shocks at the municipality level, and as noted above, $Shock_{mc}$ is the correct index.

3.2 Fixed effects related to frost shocks

The main critique in Section 2 of P-L is that KP fail to sufficiently control for confounding effects at the level of frost shocks or weather stations and that doing so reduces KP's point estimates substantially. We first revisit the reasoning behind KP's identification strategy before evaluating P-L's results.

P-L points out the key feature of KP's identification strategy, namely that KP does not use frost shocks themselves as an instrument for emigration. This is because frost shocks may have confounding effects on KP's outcomes of interest that do not go via emigration. Instead, KP exploits the interaction effect between frost shocks and port proximity for identification. This allows KP to control for the direct effect of the frost shocks. The way KP controls for frost shocks is displayed in equations 3 and 4 above, reproduced from KP. As P-L notes, in the baseline model, $Shock_{mc}$ enters linearly. However, P-L suggests that it is necessary to control for shocks in a non-linear way by including a full set of indicators for each level of frost shocks. Doing so results in first-stage point estimates that are smaller as well as more imprecise than in KP.

We discuss non-linear controls for frost shocks below, but we start by noting that $Shock_{mc}$ also enters linearly in the definition of KP's instrument, i.e. the interaction between frost shocks and proximity to port. In other words, KP uses linear frost shock variation as identifying variation. Thus, we argue that it is natural to also control for the direct effect of frost shocks linearly. Relative to using frost shocks directly as an instrument, KP propose this as a strategy that provides arguably more credible identification in terms of fulfilling the exclusion restriction. Importantly, in KP we do not claim to control for $Shock_{mc}$ in a saturated way by including fixed effects for each frost shock value. Doing so would drastically reduce the available identifying variation and was never considered as an option. Neither has it ever been proposed by any seminar participant (other than Pettersson-Lidbom), journal referee, journal editor, job market interviewer, or dissertation committee member since 2015 when we started presenting this work. We stand by the approach used in KP and have always been transparent to audiences about our identification strategy.

Nevertheless, one might wonder if frost shocks indeed have some non-linear impact on our outcomes such that it is not sufficient to control for them linearly. Therefore, KP already includes a series of robustness tests where the linearity assumption is relaxed. These are described in the robustness section of KP and displayed in Appendix Tables B.10 and B.11. The tests not only allow for polynomial functions of frost shocks (up to the 4th degree), but also for linear and cubic splines in both frost shocks and port proximity. In addition, we also control for frost shocks in logs instead of levels. KP's results are robust to all these tests for non-linear confounding effects. We note here that allowing for splines was initially proposed by Pettersson-Lidbom.

In the view of P-L, however, the above specifications are insufficient and only models saturated in all unique values of frost shocks would make identification credible. Since the frost shock variable takes on 12 distinct values and may not be sufficiently approximated by continuous functions such as polynomials, we proceed here in what we see as a natural way to control for frost shocks in a less parametric way. We specify models that include fixed effects for groups of municipalities that experience *similar* levels of frost shocks 1864–67. This will allow for any unobserved and non-linear confounding effects at the level of larger groups of municipalities, without excessively reducing statistical power.

Table 2 shows our results, with Panel A showing first-stage effects and Panel B showing reduced-form effects on labor movement participation. In column 1, we reproduce the main estimate from KP. In column 2, we include fixed effects for 3 groups of frost shocks, corresponding to municipalities with low, medium and high incidence. More specifically, we group together municipalities that experienced 0 to 3 shocks, 4 to 7 shocks, or 8 to 11 shocks, respectively. The results in both panels are essentially unchanged when allowing for arbitrary confounding effects of frost shocks at this level.

In column 3, we further restrict the comparison by creating a total of 6 groups. This specification thus controls for any unobserved effects of frost shocks among municipalities that had almost identical exposure to frost in the 1864–67 period, i.e. either 0 to 1 shocks, 2 to 3 shocks, 4 to 5 shocks, and so on. The results are again similar to the main KP results in column 1. Hence, even within such narrowly defined groups, fixed effects produce statistically and economically significant results.

For reference, P-L’s estimates when using the full set of frost shock fixed effects are shown in column 4. While the reduced-form effect in panel B is similar to KP, the first-stage point estimate becomes both markedly smaller and more imprecise when using P-L’s specification as compared to the main specification in column 1. P-L’s estimate reduces the KP point estimate by almost 62 percent. This stands in sharp contrast to the case of including 6 frost shock groups, which reduces the point estimate by only 8 percent.¹⁵ While a clear argument for why a saturated model is needed is lacking in P-L, we do not argue that a model with 3 or 6 groups is necessarily more reasonable. However, in our view, one must hold quite stark beliefs about the nature of frost shocks’ confounding effects if they should only become noticeable once controlling for the *precise* level of frost shocks, rather than pairs of shock levels.

P-L’s related suggestion to include weather-station fixed effects is similar but even more demanding than including frost shock fixed effects. This is because, as we use cross-sectional variation, there is (almost) no variation in frost shocks within a weather station. Thus, not

¹⁵The second stage IV estimate with 6 frost shock groups is 0.027, similar to the baseline KP estimate of 0.023. The small first-stage estimate in P-L makes the IV estimate larger (0.066).

only does this specification require that comparisons be made with other municipalities that had the same number of frost shocks, the comparison municipalities must also be associated with the same weather station. Moreover, since port proximity is used to capture the internal cost of emigrating, estimating the effect within the more narrowly defined weather-stations likely allows for less meaningful variation in the cost of emigrating.

In column 5 of Table 2, we display results when adding weather-station fixed effects to our baseline regression. Although the point estimates decrease with this inclusion, they are also more imprecise as reflected by the increase in standard errors.¹⁶ We note that one cannot reject that point estimates with weather-station fixed effects are the same as in KP’s specifications. For instance, testing a null hypothesis that the coefficient in column 5, Panel A of Table 2 is the same as the coefficient in column 1 yields a p-value of 0.7.

3.3 Regional fixed effects and an “illustrative example”

We control for county fixed effects in order to not disproportionately rely on regional differences. However, P-L suggests that “the county level seems more or less arbitrary since these effects are only partly related to the other key control variable *Port*, which measures the geographical distance to ports”. While the administrative level chosen for a region fixed effect – the Swedish county – is arguably natural for the historical setting, one could perhaps argue for a fixed effect at another regional level, such as the three traditional lands of Sweden (*Svealand*, *Götaland*, and *Norrland*), or none at all and simply rely on the geographical controls.

In fact, Table 3 documents that the results in KP are robust to these variations in regional fixed effects. Column 1 displays the baseline specification in KP including county fixed effects. Column 2 displays fixed effect for the three lands of Sweden. Following the suggestion in P-L, columns 3 and 4 additionally display fixed effects for areas arranged into deciles or ventiles, respectively, in terms of their distance to the emigration ports.¹⁷ Finally, column 5 display results without any fixed effect.

Results are similar throughout, although the first-stage coefficient is somewhat larger and the second-stage coefficient is somewhat smaller in the alternative specifications as compared to the baseline specification in KP. Note that the result in column 5 is at odds with the, so-called, “illustrative example” displayed in Table 4 (column 1) in P-L, which suggests that our results are not robust when dropping county fixed effects from the regression model. How come? This is explained by the fact that, although KP’s identification strategy is specified

¹⁶Standard errors increase also when using alternative methods without clustering at the level of weather-stations, such as Conley’s spatial-correlation robust standard errors.

¹⁷P-L argues that “a much better approach is to order all 2,359 municipalities based on the distance to ports and define geographical groups accordingly. In this way, it is possible to control much more convincingly for the factors related to distance to ports that seem to concern KP.”

throughout with the inclusion of a set of geographical controls, P-L strips the regression model from all controls except frost shocks, proximity to port, and baseline population. Thus, the results in KP are robust to dropping regional fixed effects as long as we use our preferred specification where we compare observations that are similar on baseline observables.

4 Concluding remarks

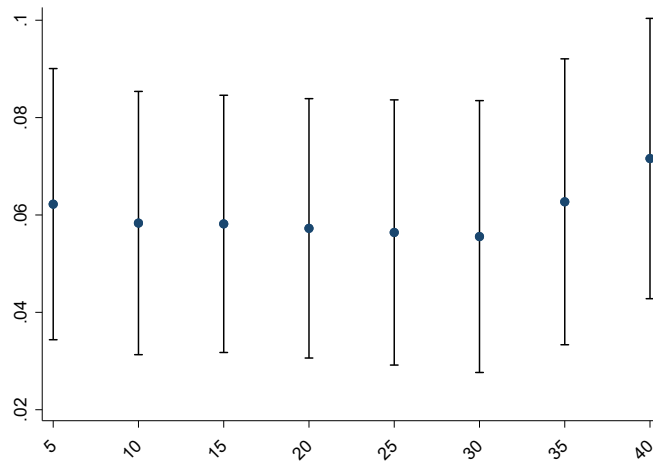
In this note, we have evaluated P-L's claims regarding the results in KP. In particular, P-L raises concerns regarding KP's emigration data and how we control for the direct effects of frost shocks.

We regret that some of the misunderstandings in his comment could not be solved by a fruitful discussion between us. Yet we are glad to have discussed some aspects here that we did not make room for in the original paper. Although P-L raises some interesting matters, which in part point at the limits of our identifying variation, we find that our results stand robust to his criticism. Measurement error does not appear to have a significant impact on our results and both this reply and KP itself show that we are able to address concerns about non-linear confounders in a wide range of specifications.

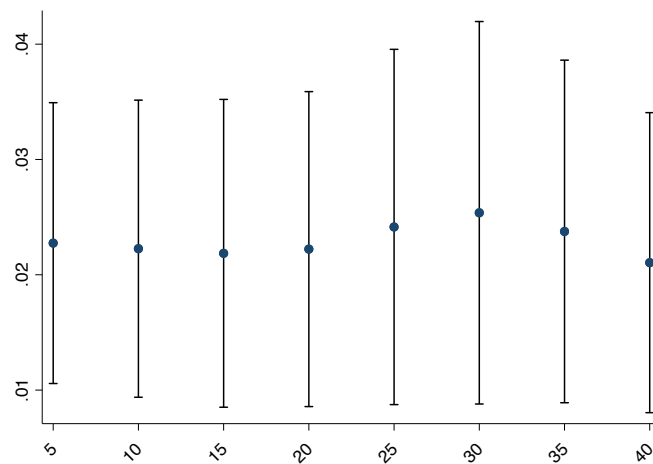
References

- Ahlqvist, G. (1976). Sydsvensk utvandring till Danmark. fallstudie av en i kyrkoböckerna oregistrerad utvandring. *Scandia*, 42(2), 193–206.
- Bohlin, J. & Eurenus, A.-M. (2010). Why they moved—emigration from the Swedish countryside to the United States, 1881–1910. *Explorations in Economic History*, 47(4), 533–551.
- Clemensson, P. (1996). Göteborgska källor till den stora utvandringen. *Dokumentet*, (3).
- Eriksson, I. (1969). Passenger lists and the annual parish reports on emigrants as sources for the study of emigration from Sweden. Technical report, Nordic Emigration: Research Conference in Uppsala Sept. 1969.
- Johansson, R. (1976). Registrering av flyttare, en källkritisk granskning av kyrkoboksmaterial 1840–90. *Scandia*, 4(2), 167–192.
- Karadja, M. & Prawitz, E. (2019). Exit, voice and political change: Evidence from Swedish mass migration to the United States. *Journal of Political Economy*, 127(4).
- Kronborg, B. & Nilsson, T. (1975). *Stadsflyttare : industrialisering, migration och social mobilitet med utgångspunkt från Halmstad 1870-1910*. PhD thesis, Uppsala University, Faculty of Arts. Diss. av båda förf.
- Ljungberg, J. (1997). The impact of the great emigration on the Swedish economy. *Scandinavian Economic History Review*, 45(2), 159–189.
- Norman, H. (1974). *Från Bergslagen till Nordamerika : studier i migrationsmönster, social rörlighet och demografisk struktur med utgångspunkt från Örebro län 1851-1915*. PhD thesis, Uppsala University, Faculty of Arts.
- Odén, B. (1964). Den urbana emigrationen från Sverige 1840-1872. Technical report, Unpublished manuscript, Lund.
- Odén, B. (1971). Ekonomiska emigrationsmodeller och historisk forskning. ett diskussionssinlägg. *Scandia*, 37(1), 1–70.
- Pettersson-Lidbom, P. (2020). Exit, voice and political change: Evidence from Swedish mass migration to the United States a comment. Technical report.
- Stephens, M. & Unayama, T. (2019). Estimating the Impacts of Program Benefits: Using Instrumental Variables with Underreported and Imputed Data. *The Review of Economics and Statistics*, 101(3), 468–475.

- Sundbärg, G. (Ed.). (1913). *Emigrationsutredningen: Betänkande*. Norstedt & Söner.
- Tedebrand, L.-G. (1972). *Västernorrland och Nordamerika 1875-1913 : utvandring och återinvandring*. PhD thesis, Uppsala University, Faculty of Arts.
- Tedebrand, L.-G. (1976). Sources for the history of Swedish emigration. In H. Runblom & H. Norman (Eds.), *From Sweden to America. A history of migration*. Uppsala: Acta Universitatis Upsaliensis, pp. 76-93.
- Verneresson Wiberg, A.-K. (2016). Migration och identitet: En studie av utvandringen från Blekinge till Danmark och Tyskland 1860–1914. Technical report, Uppsala: Department of Social and Economic Geography, Uppsala University.
- Westerståhl, J. (1945). Svensk fackföreningsrörelse: organisationsproblem, verksamhetsformer, förhållande till staten.



(a) Emigration (OLS)



(b) Labor movement (IV)

Figure 1

Test of measurement error due to European emigration

Notes: This figure documents the first-stage relationship between the instrument and log emigrants in (panel A) and the second-stage relationship between log emigrants and labor movement participation share (panel B), for separate samples depending on the latitude of the municipality. Each value i on the x-axis denotes that the regression sample does not include municipalities up till the i th percentile, $i = 5, 15, 20, \dots, 40$. Included controls are county fixed effects, frost shocks 1864–1867, proximity to the nearest emigration port, nearest trade port, nearest weather station, nearest town and Stockholm, log population in 1865, log area, latitude, longitude, arable land share in 1810 and indicators for urban municipalities and high soil suitability for the production of barley, oats, wheat, dairy and timber as well as the interaction between growing season frost shocks and proximity to the nearest town and trade port, respectively. Standard errors are clustered at the weather station level. Bars around point estimates represent 95 percent confidence intervals.

Table 1 Emigration using different sources and counted over different periods

	KP 1867–1920	KP 1867–1895	Church books 1867–1895	Passenger 1869–1920	KP (inflated) 1867–1920	KP 1867–1884	KP 1885–1893
<i>Panel A</i>							
Dependent variable:	Emigration						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Shocks \times Proximity to port	0.0632*** (0.0135) [0.0357,0.0907]	0.0672*** (0.0161) [0.0344,0.1001]	0.0635*** (0.0177) [0.0275,0.0995]	0.0429*** (0.0147) [0.0131,0.0728]	0.0637*** (0.0136) [0.0360,0.0913]	0.0731*** (0.0204) [0.0315,0.1147]	0.0609*** (0.0127) [0.0351,0.0867]
<i>Panel B</i>							
Dependent variable:	Labor Movement						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Emigration	0.0234*** (0.0063) [0.0111,0.0357]	0.0220*** (0.0065) [0.0092,0.0348]	0.0233*** (0.0074) [0.0087,0.0378]	0.0345*** (0.0115) [0.0120,0.0569]	0.0233*** (0.0062) [0.0110,0.0355]	0.0202*** (0.0067) [0.0071,0.0332]	0.0244*** (0.0065) [0.0116,0.0372]
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Shocks \times Market Access	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2358	2358	2358	2358	2358	2358	2358
F-statistic	21.99	17.50	12.99	8.63	22.01	12.99	22.93

Notes: OLS and IV regressions. This table documents the first-stage and second-stage relationships using different sources and counting emigrants over different periods. Emigration is defined as the log number of emigrants. Columns 1 and 2 display the results using the emigration variable from KP for the period 1867–1920 and 1867–1895, respectively. Column 3 and 4 display the results using only the emigrant lists from the church books (1867–1895) and the passenger lists (1869–1920), respectively. Column 5 inflates the KP emigration numbers (1867–1920) by a factor 1.3699 (assuming $\theta = 0.73$). Columns 6 and 7 use the emigration variable in KP for the years 1867–1884 or 1885–1893, respectively. *County fixed effects* denotes fixed effects at the county level. *Controls* denotes the inclusion of the following control variables: growing season frost shocks 1864–1867, proximity to the nearest emigration port, nearest town, nearest trade port, nearest weather station and Stockholm, log of the population at baseline, log area, latitude, longitude, as well as an urban indicator and a set of indicators for high soil quality for the production of barley, oats, wheat, dairy and timber. *Shocks \times Market Access* denotes the inclusion of the interaction between growing season frost shocks and proximity to the nearest town and trade port, respectively. The F-statistic refers to the excluded instrument. Standard errors (in parenthesis) are clustered at the weather-station level. 95 percent confidence intervals in brackets. P-values are denoted by the following scheme: *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table 2 Adding frost-shock or weather-station fixed effects

	KP baseline	3 frost groups	6 frost groups	12 frost groups	Weather-station fe
<i>Panel A</i>					
Dependent variable:	Emigration				
	(1)	(2)	(3)	(4)	(5)
Shocks \times Proximity to port	0.0632*** (0.0135) [0.0357,0.0907]	0.0661*** (0.0133) [0.0389,0.0933]	0.0582*** (0.0155) [0.0265,0.0899]	0.0243 (0.0193) [-0.0149,0.0636]	0.0341 (0.0367) [-0.0408,0.1090]
<i>Panel B</i>					
Dependent variable:	Labor Movement				
	(1)	(2)	(3)	(4)	(5)
Shocks \times Proximity to port	0.0015*** (0.0003) [0.0008,0.0021]	0.0016*** (0.0004) [0.0009,0.0024]	0.0016*** (0.0004) [0.0008,0.0024]	0.0016** (0.0008) [0.0000,0.0033]	0.0012 (0.0009) [-0.0006,0.0030]
County fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Shocks \times Market Access	Yes	Yes	Yes	Yes	Yes
Frost shock fixed effects	No	Yes	Yes	Yes	No
Weather station fixed effects	No	No	No	No	Yes
Observations	2358	2358	2358	2358	2358

Notes: OLS regressions. This table documents the first-stage relationship between log emigration and the instrument in panel A and the reduced-form relationship between labor movement participation share and the instrument in panel B. *Frost shock fixed effects* denotes the inclusion of fixed effects for different groups of frost-shock values. *Weather-station fixed effects* denotes the inclusion of fixed effects at the weather-station level. *County fixed effects* denotes fixed effects at the county level. *Controls* denotes the inclusion of the following control variables: growing season frost shocks 1864–1867, proximity to the nearest emigration port, nearest town, nearest trade port, nearest weather station and Stockholm, log of the population at baseline, log area, latitude, longitude, as well as an urban indicator and a set of indicators for high soil quality for the production of barley, oats, wheat, dairy and timber. *Shocks \times Market Access* denotes the inclusion of the interaction between growing season frost shocks and proximity to the nearest town and trade port, respectively. Standard errors (in parenthesis) are clustered at the weather-station level. 95 percent confidence intervals in brackets. P-values are denoted by the following scheme: *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table 3 Altering or dropping regional fixed effects

	County	Major regions	Deciles	Ventiles	None
<i>Panel A</i>					
Dependent variable:	Emigration				
	(1)	(2)	(3)	(4)	(5)
Shocks \times Proximity to port	0.0632*** (0.0135) [0.0357,0.0907]	0.0910*** (0.0119) [0.0668,0.1152]	0.0887*** (0.0181) [0.0517,0.1256]	0.0822*** (0.0185) [0.0445,0.1199]	0.1030*** (0.0158) [0.0707,0.1354]
<i>Panel B</i>					
Dependent variable:	Labor Movement				
	(1)	(2)	(3)	(4)	(5)
Shocks \times Proximity to port	0.0015*** (0.0003) [0.0008,0.0021]	0.0013*** (0.0003) [0.0006,0.0020]	0.0012*** (0.0003) [0.0006,0.0018]	0.0011*** (0.0003) [0.0005,0.0017]	0.0009** (0.0003) [0.0002,0.0016]
Controls	Yes	Yes	Yes	Yes	Yes
Shocks \times Market Access	Yes	Yes	Yes	Yes	Yes
Observations	2358	2358	2358	2358	2358

Notes: OLS regressions. This table documents the first-stage relationship between log emigration and the instrument in Panel A and the reduced-form relationship between labor movement participation share and the instrument in Panel B. Column 1 displays the baseline specification in KP using county fixed effects. Column 2 displays fixed effect for the three lands of Sweden. Column 3 and 4 display fixed effects for areas arranged into deciles or ventiles, respectively, in terms of their distance to the emigration ports. Column 5 display results without any fixed effect. *Controls* denotes the inclusion of the following control variables: growing season frost shocks 1864–1867, proximity to the nearest emigration port, nearest town, nearest trade port, nearest weather station and Stockholm, log of the population at baseline, log area, latitude, longitude, as well as an urban indicator and a set of indicators for high soil quality for the production of barley, oats, wheat, dairy and timber. *Shocks \times Market Access* denotes the inclusion of the interaction between growing season frost shocks and proximity to the nearest town and trade port, respectively. Standard errors (in parenthesis) are clustered at the weather-station level. 95 percent confidence intervals in brackets. P-values are denoted by the following scheme: *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.