Robert Lange
student no. 5632196
examination no. 33252

# The Impact of Skilled-Worker Immigration on the Diffusion of New Technologies

Bachelor Thesis

Supervisor: Prof. Dr. Susanne Prantl

Submitted for the Bachelor examination in

Economics

Faculty of Management, Economics and Social Sciences
University of Cologne

Cologne, 2016

# Contents

# List of Figures



# List of Tables



# Symbol directory

Symbols and abbreviations refer to the following, except where stated differently.

| | |
|---|---|
| IV | Instrumental Variable Regression |
| MCAR | Missingness Completely at Random |
| MAR | Missingness at Random |
| OLS | Ordinary Least Squares |
| OVB | Omitted Variable Bias |
| $\theta$ | Parameter of Interest |

# 1 Introduction

The impact of immigration has never been as important as today. Ever since the tragic events of Paris in 2015, the refugee crisis in Europe, caused by geopolitic conflicts all around the world, and the dramatic happenings of New Years' Eve in Cologne, Stuttgart and Hamburg a mostly normative and populist discussion about the effects of immigration has emerged. Right-wing movements all across Europe and the emergence of politicians such as Donald Trump and Marine Le Pen highlight the severe emotional intensity of this topic. The aim of this thesis is to turn away from the public discourse and to analyze a very specific question in a positive way: How does the immigration of a selection of highly skilled workers affect the diffusion of new ideas and technologies? Technology diffusion is defined as "(...) the process by which an innovation is communicated over time, through certain channels, among the members of a social system." (Williams and Baláž, 2008, 73). In order to fully understand the implications of migration, it is crucial to grasp the mechanisms that determine the transition of the adoption process from the short to the long-run. Therefore, the following pages are going to empirically and theoretically examine the question at hand.

At the center of this analysis is going to be the critical examination of the study conducted by Erik Hornung (2014a). In his unique study Hornung tries to enlighten the discussion by focussing on knowledge spillovers created by high-skill immigration. The author concentrates on the immigration of French Protestants to Prussia during the 16th and 17th century. His regression models indicate that there is a positive effect of immigrants' population share on the firm-level production output. By using this very special French historical setting Hornung (2014a, 85) is able to circumvent distortionary effects of technology diffusion caused by information channels in today's information age. Consequently, he is able to argue that most of the communication at that time was done by face-to-face communication (Hornung, 2014a, 88). This is the necessary condition for his assumption that a good proxy for this single diffusion channel can be found in the population share of Huguenots (Hornung, 2014a, 100).

The main ambition of this thesis is to highlight major problems in the Hornung's study and to emphasize the differences between short and long-run implications of immigration in the market for ideas and knowledge. I draw attention to several cultural, institutional and sociological aspects of high-skill migration such as native's attitude towards immigrants, "peer effects" observed at the academic level and social networks, which can foster integration and assimilation of human capital. I find it important to state that even in already highly skilled groups certain exceptional individuals, so-called knowledge brokers (Williams and Baláž, 2008, 75-76), often times seem to drive technology developments.

The remainder of this thesis is structured as follows. Chapter 2 introduces the reader to the analysis of Hornung. The historical events that led to the immigration of 43,000

Huguenots to Germany (Lausberg, 2007, 73) are summarized. An overview of benefits and challenges Huguenots faced in Brandenburg-Prussia is given. Afterwards, the data set constructed by Hornung is described. Furthermore, the theory behind Hornung's identification strategy is reviewed. I display the regression models used and state the driving identification assumptions behind instrumental variables (IV) estimation. In the end, some specific regression results are analyzed. Chapter 3 turns the attention towards potential problems regarding key assumptions like a failure of instrument exogeneity and weak instrument consequences. I question Hornung's research design with regard to small sample size, little exogenous variation in the instrument and assumed missing data mechanism. I highlight the potential and dangers of this and compute robustness checks using different imputation algorithms. Furthermore, I take a deeper look at the inference conducted by Hornung and estimate standard errors using a superior method: clustered bootstrap standard errors. I connect both topics by examining the importance of imputation for inferential results. Chapter 4 relates the results established by Hornung with further important literature results regarding high-skill immigration and technology diffusion. The studies conducted by Waldinger (2010), Moser, Voena and Waldinger (2014) and Borjas and Doran (2012) emphasize different versions of one special trade-off: knowledge spillovers versus resource scarcity. Additionally, I deal with the question how the speed of adjustment, incentives through networks, and language adoption affect the diffusion of technology. The fifth chapter summarizes and concludes.

# 2 Hornung's Analysis (2014): Huguenot Immigration and Technology Diffusion

The following chapter illustrates the analysis by Erik Hornung (2014a). First, the historical events of the 17th and 18th century, that led to many Huguenots conceding their long-time homes, are reviewed. Afterwards, the data set compiled by Hornung is studied. Furthermore, I analyze and interpret the identification strategy used to estimate the parameter of interest and the underlying assumptions. Afterwards, I carry out a first attempt to evaluate the qualitative effect and the quantitative strength of Hornung's results.

## 2.1 Huguenot Immigration to Prussia: A Historical Review

In the center of Erik Hornung's analysis stand the events that led to the immigration of 20,000 Huguenots to Brandenburg-Prussia. The following section outlines the history of persecution of Huguenots and their eventual settlement in Prussia.

The beginnings of the reformation in France were located in Paris in the 15th century. After the university Paris-Sorbonne dismissed the teaching of Martin Luther in 1521 and the Lutheran and Augustinian monk Jean Valliere was burned in 1523, the persecution

of French protestants began (Lausberg, 2007, 15). Despite this cruelty, the reformation spread in almost all provinces in France. The most important protagonists were Guillaume Farel (1489-1565) and Jean Calvin (1509-1564) who gave the movement a theological and organizational unified system (Lausberg, 2007, 17). After several years of pursuit and negotiation, persecution reached its zenith in the "Massacre de la Saint-Barthélemy" (1572) and the following Huguenot Wars (1562-1598) (Hornung, 2014a, 90). Heinrich IV created the Edict of Nantes in 1598 which granted reformed French Protestants religious freedom. It was the first hope for peaceful integration of French Protestants. Estimates state that in 1661 approximately one million Huguenots lived in France and were part of all social classes (Lausberg, 2007, 48). In the 17th century King Louis XIV of France started to doubt the loyalty of French Protestants and again rejected Protestantism (Lausberg, 2007, 49). After Louis XIV revoked the Edict of Nantes in 1685 more than 200,000 Huguenots decided to flee in the hope of a better life (Hornung, 2014a, 90). 43,000 left France in order to come to Germany and 20,000 refugees decided to move to Prussia (Hornung, 2014a, 90; Lausberg, 2007, 73). Therefore, Brandenburg-Prussia is often called the most important place of refuge in Germany during this time (Lausberg, 2007, 73). This was mainly due to important structural admission requirements and benefits granted to the newcomers in the Edict of Potsdam (Lausberg, 2007, 73).
In the Edict of Potsdam, which came into power in October 1685, Frederick William, the Great Elector of Brandenburg (*"Kurfürst"*), constituted 14 articles which determined the general conditions for the Huguenot settlement (Lausberg, 2007, 73). The Edict promised help during the flight and support for immigration and settlement. The autochthonous population in Brandenburg-Prussia had to share their food with the new arrivals (Lausberg, 2007, 74). Furthermore, the Great Elector of Brandenburg offered dilapidated and abandoned houses and the necessary materials to reassemble these (Lausberg, 2007, 74). The Huguenots had to pay no taxes and got support (financial payments and privileges) for establishing new manufactories (Lausberg, 2007, 74). They were also allowed to practice their reformed religion in French and to nominate own priests (Lausberg, 2007, 74). Lausberg (2007, 195) states that these benevolent benefits were due to a mix of power-political goals, economic interests and the necessity to help fellow-believers. Huguenots were known as being loyal and, therefore, could stabilize the absolutist system (Lausberg, 2007, 195). Furthermore, approximately 500 Huguenot soldiers joined the Prussian military and the Huguenots were famous for their well-educated manufacturers, which revived the German economy after the Thirty Years' War (Lausberg, 2007, 195). Therefore, the reasons for admitting the Huguenot immigrants seem to be manifold.

Because Hornung (2014a, 95) states, that the settlement pattern of Huguenot immigrants was mainly determined by Prussian rulers, it is important to know more about this aspect: The Edict of Potsdam left the specific town choice up to the immigrants, but recommendations were made (Lausberg, 2007, 74). This could have been due to the fact that many towns were still destroyed and incapable of providing the necessary food for the population. Hornung (2014a, 95) assumes that Calvinist immigrants faced "(...) a situation where an absolutist bureaucracy controlled the settlement of Huguenots and determined the scope and the direction of their economic activities.". He concludes that the settlement decision of Huguenots was not only endogenously determined by immigrant preferences but that they were assigned to towns that needed the economic support most. Frederick Wilhelm thereby tried to repopulate regions that were almost vanished by the Black Death and the Thirty Years' War (1618-1648) (Hornung, 2014a, 96).

Lausberg (2007, 200), on the other hand, finds several other sources of different settlement patterns. He states that many French Protestants placed importance on settling in towns with good economic conditions. For example, many Huguenots established themselves in the town of Erlangen because it had flourishing trade connections with England and the Netherlands (Lausberg, 2007, 200). It was also well-known that the university in Frankfurt/Oder had a great reputation and attracted many Calvinist intellectuals (Lausberg, 2007, 200). The Great Elector encouraged this development by granting scholarships (Lausberg, 2007, 200). Furthermore many Huguenots worked in the military before they had to leave France. During this time Frederick William deployed many regiments in Prussia and, therefore, many refugees went on to settle there (Lausberg, 2007, 200).

Although the Edict of Potsdam stipulated the benevolent treatment of the immigrants to the native population, there are some reports of aloofness and even open disapproval (Lausberg, 2007, 207). Because the French Protestants settled in established towns and communities, they had to come to terms with different cultural norms and values (Lausberg, 2007, 207). Many natives were jealous of the privileges granted to Huguenots and were anxious about their own professional careers (Lausberg, 2007, 208). But not only jealousness and increased competition were reasons for open disapproval. The literature also reports that German natives were partly just racist and intolerant towards customs and traditions of the Huguenots (Lausberg, 2007, 209). Sometimes they even had to face violence (Lausberg, 2007, 209). Vocational training contracts (*"Ausbildungsverträge"*) led to a clear improvement (Lausberg, 2007, 211). Many Germans saw the great technological lead of the Huguenots and decided to sent their children into apprenticeship. This caused not only mutual exchange and diffusion of knowledge, but also the revocation of stereotypes and intolerance (Lausberg, 2007, 211). Furthermore, marriages between Huguenots and natives increased from the beginning of the 18th century (Lausberg, 2007, 211). This further fostered the cultural assimilation into the Prussian society. All in all, Lausberg (2007, 210) concludes that Huguenots were accepted when the established population received clear economic or cultural benefits from their migration.

## 2.2 Exploited Data Set: Firm-level, Immigration Share and Population Loss

In his empirical analysis Hornung combines multiple sources of data into one unique data set in order to estimate the long-term effect of high-skill immigration on the overall economic output through technology assimilation. The data can be partitioned into three main categories: The firm-level data on productivity of Prussian manufactories in 1802, the data on immigrants living in Brandenburg-Prussia, and the population loss data from the Thirty Years' War, which is needed to construct an instrument candidate (Hornung, 2014a, 96 et sqq.).

The firm-level data is constructed by using the "Register of Factories in the Prussian State" from the Royal Prussian Privy Filing Department in 1802 taken from Krug (1805) (qtd. in (Hornung, 2014a, 96)). The manufactories had to fill out a standardized survey, which contained information on their classification, the value of the raw input materials and the value of production (Hornung, 2014c, 2). Nominal valued variables are all measured in Prussian Thalers. All factories that were established within the Prussian borders by 1802 are included except for those located in Ansbach, Bayreuth, Neuchatel, Silesia, and those which were located in areas that were gained as compensation for losses in the French war (Hornung, 2014a, 97). The original data set includes data on 750 textiles and 1025 non-textiles manufactories. After eliminating all manufactories in rural parts of Brandenburg-Prussia and all manufactories for which there is no information available in 1807 the data set shrinks down to 693 textiles and 694 non-textiles manufactories (Hornung, 2014c, 2). One major problem that arises from using data from the 18th century is that 15.2 and 23.8 percent of the data for input values from textiles and non-textiles manufactories are missing, respectively. Hornung (2014c, 2) states that the mechanism behind this data problem seems to be depending on the location of the manufactories. In accordance with Cameron and Trivedi (2009, 47), dropping such a big portion of the dataset could lead to selection bias if the carry over is dependent on unobservable variables. In order to preserve the sample size the author makes use of imputation methods. The missing data is substituted with imputed data from a linear regression for continuous variables. Hornung (2014c, 2), thereby, uses univariate multiple imputation methods.

Huguenot immigration data is taken from the "Rôle général des Français refugiez dans les Estats de la Majesté le Roy de Prusse", which is an annual Prussian register in which French protestant had to document their personal data (Hornung, 2014a, 98).[1] Due to continuous fluctuations in the first years after the Edict of Potsdam, the author decided to generalize the effect of immigration to the first generation (Hornung, 2014a, 98). Therefore, Huguenot population data is taken for the year 1700 instead of 1685. Due to data problems the Prussian share of Huguenots is defined as the number of Huguenots living

[1] In order to compute the population share quotient overall town population data is taken from Schmoller (1922, 272-284) (qtd. in Hornung (2014a, 98)).

in Prussian towns in 1700 divided by the total town population in 1730 (Hornung, 2014a, 98).

Finally, Hornung constructs an instrumental variable candidate by arguing that even though the settlement pattern of Huguenot immigrants might not have been fully exogenous, it might be possible to extract the exogenous variation/information by making use of the fact that Frederick Wilhelm deterministically orchestrated the immigration flows. It is often stated throughout the historiography that Huguenots were directly advised to settle in towns that were weakened by the occurrence of the Black Death and the Thirty Years' War. The author, therefore, argues that the endogenous Huguenot population share can be instrumented by the population loss that towns experienced (Hornung, 2014a, 100). In order to come up with credible population loss values on the town level, Hornung makes again use of multiple data sources: The Handbook of German Towns (Keyser, 1939-1974), Behre (1905), and Wohlfeil (1976) (qtd. in Hornung (2014a, 101)). Since Brandenburg started to participate in the Thirty Years' War in 1626, population data is needed for 1625 and 1652 (Hornung, 2014c, 3). But there is no data for all towns available from Keyser (1939-1974) (qtd. in Hornung (2014a, 101)). Hornung (2014c, 3) combats this problem by interpolating the latest population data in the period with German population growth rates taken from Pfister (2007, 10) (qtd. in Hornung (2014c, 3)). In my opinion, this introduces a general problem of interpolation, namely location-mismatch: interpolating town data with country growth rates. If Prussian towns did not grow at the same averaged German rate this could be a possible source of bias. Hornung (2014c, 4) selects towns upon a number of different reasonable criteria such as consistency of data reports (information before and after war, same unit of observation). As a result, only 57 towns remain in the studied sample. But this small sample size can cause suspicion. In order to construct more "robust" data Hornung (2014a, 101) combines all three instrument sources by averaging. A problem that remains is that if all sources are weak than combining them is not going to increase their informational content.

## 2.3 Applied Identification Strategies: OLS and IV

This section introduces the reader to the two main identification strategies employed by Hornung in order to estimate the causal effect of first-generation Huguenot immigration on economic output in Brandenburg-Prussia. Examining figures 1 and 2 reveals that there is a clear positive correlation between the observed town population share of Huguenots in 1700 and the economic output produced by firms in this town in 1802. Furthermore the effect seems to be larger for the textile industry than for non-textile industries. The main aim of Hornung is to prove that this observed correlation is due to a causal relationship. As a first approach he controls for several other effects that could have influenced the observed output measure in simple ordinary least squares (OLS) regressions (Hornung, 2014a, 104). It will become apparent that there remain several problems such as selection bias, omitted variable bias (OVB) and measurement error that could poison the estimator. All of those would lead to the endogeneity of the observed population share of Huguenots. Hornung (2014a, 100) proposes to only make use of the exogenous variation in the population share that was introduced by the population loss towns suffered during the Thirty Years' War. He argues that Frederick Wilhelm orchestrated the geographic settlement by recommending that immigrants should settle in towns that had experienced great loss (Hornung, 2014a, 101). Using the exogenous experience of death during this time he comes up with an instrumental variable regression (IV) approach, which estimates a positive effect of the presence of first generation immigrants on the economic output observed 100 years after the first wave of migration.

### 2.3.1 Ordinary Least Squares Estimation

The OLS model of Hornung's analysis estimates the following logarithmic form of a Cobb-Douglas production function (Hornung, 2014a, 98):

$$y_{ij} = \alpha + \beta_l l_{ij} + \beta_k k_{ij} + \beta_m m_{ij} + \theta \left( \frac{H_{1700}}{TP_{1730}} \right)_j + X_j' \gamma + \varepsilon_{ij}. \tag{1}$$

The dependent variable $y_{ij}$ is the natural logarithm of the total output of a firm $i$ in town $j$ in 1802 measured in Prussian thalers. $l_{ij}$ denotes the natural logarithm of the number of workers, $k_{ij}$ the natural logarithm of the number of looms used in production and $m_{ij}$ the natural logarithm of the value of input materials used in the production process in 1802 (Hornung, 2014a, 99). In order to control for several town and economic conditions an additional vector of controls, $X_j$, is included (Hornung, 2014a, 99). Some of those are the overall town population in 1802, the percentage of Protestants living in the town in 1826, measures of access to production materials and whether the town had a relevant textile production in 1680 (Hornung, 2014a, 99). The regressor of interest is $(H/TP)_j$ which represents the quotient of Huguenot individuals living in town $j$ in 1700 divided by the total population of town $j$ in 1730. The workforce of French Protestants is therefore not directly included into the production process of firm $i$. It is likely that social and professional interaction between the autochthonous population and Huguenots was higher in towns with a bigger population share of Huguenots (Hornung, 2014a, 100). Hornung (2014a, 107), therefore, argues that the hypothesis of Huguenots increasing production output through diffusing knowledge by their presence in the Prussian communities can be tested.

Hornung (2014a, 104) estimates several specifications with different controls, which I partly report in table 1. Column 1 shows that there is a highly significant positive relationship/correlation between the regressand, the natural logarithm of firm output in 1802, and the regressor, the population share of French Protestants in 1700 in the firm's town. The estimated parameter of interest, $\hat{\theta}$, can be interpreted as the percentage response in firm output as the population share of Huguenots increases by one percent (Hornung, 2014a, 103). Therefore, the bivariate regression suggests that economic output rose on average 10 percent if the population share of first-generation Huguenots increased by one percent, *ceteris paribus*. Adding several control variables significantly decreases the magnitude of this estimate. When adding several production input and town specific variables one finds that the economic output only seems to have increased on average and c.p. by 1.4 percent (see table 1 column 3). This estimator remains largely unchanged when further controlling for town and religion specific variables such as population and the share of protestants in the county. The most significant decrease in the estimated value can be found when controlling for the number of workers occupied in the firms (see table 1 column 2).[2] This is no surprise because it seems likely that Prussian firms also employed Huguenot workers directly in their manufactories. Huguenots were on average more educated than Prussian workers and hence also more productive in general (Lausberg, 2007, 195). Controlling for the otherwise omitted variable is absorbing a large portion of the otherwise "contaminated" effect. Hornung (2014a, 103) estimates cluster-robust standard errors on the town level because the data may be clustered, with observations correlated within towns and independent across towns. In all OLS specifications the coefficient is highly significant up to the one percent level. Therefore, the null hypothesis of insignificance ($H_0 : \theta = 0$) can be rejected for any common level.

The coefficient of determination, $R^2$, increases dramatically when controlling for production inputs and further town specific variables. After extending the estimated regression equation, around 96 percent of the total variation in the observed data is explained by the model. Hence, this goodness-of-fit measure suggests a very good fit of the model at hand. But this conclusion might be deceiving. The $R^2$ always increases when adding more explanatory variables. Factoring this in, one can instead compute the adjusted $R^2$. The models displayed in table 1 do not suffer when doing so. This is due to the fact that almost all regressors have large t-statistics (Wooldridge, 2012, 202). All in all, first baseline OLS regressions predict a positive relationship between first-generation Huguenot immigration to Prussia and the economic output 100 years later.

### 2.3.2 Instrumental Variables Estimation

Concluding that there is causal relation between the two variables of interest could be rash. It is possible that immigrants who settled in Prussia were very different from the ones that settled in Hessen for example. If this selection was positive and Prussian Huguenots were on average more productive and technology sharing then OLS estimates would be biased. As a result, the true effect would be overestimated. Furthermore, there are two different potential sources of bias. It is possible that the base OLS specification does not control for variables that are correlated with both the explained and the explanatory

[2] Regressions can be reproduced with the supplementary code provided with this thesis.

variable. Huguenots might have moved to towns that had already established French or Protestant communities before 1700. If this community measure had a significant influence on economic performance than we would suffer from OVB. Hornung's data does not offer information on whether there were already established networks, Huguenots could use before 1700. Such omitted variables would imply that the exogeneity assumption, needed to consistently estimate an unbiased effect, would fail. Increasing the sample size would not help to overcome this problem. For such cases Wooldridge (2010, 89) states that not only the estimator of the parameter of interest, but also all other estimators are going to be biased. One also has to be worried about measurement error introduced by missing values, merging data sets and the collection of data in the 18th and 19th century. Measurement error in the regressors would lead to the same failure of the exogeneity assumption, whereas measurement error in the regressand would "only" introduce greater variance in the estimator (Dougherty, 2011, 308 et sqq.).
One common technique to solve all the mentioned problems is to find a valid instrument for the possible endogenous regressor, $(H/TP)_j$ (Angrist and Krueger, 2001, 71 et sqq.; Verbeek, 2008, 131 et sqq.). A valid instrument, $z$, has to fulfill the following assumptions:

**IV Assumption 1** *Instrument Relevance*

The instrument has to be partially correlated with the endogenous regressor after controlling for all other exogenous regressors (Wooldridge, 2010, 90).

$$Cov\left(z, \left(\frac{H}{TP}\right)_j\right) \neq 0. \tag{2}$$

Hence this is a stronger assumption than just correlation between the instrument and the "poisoned" regressor implied by equation (2). It describes that some of the exogenous variation in the endogenous variable $(H/TP)_j$ has to be captured in the instrument after excluding influences caused by other regressors.

**IV Assumption 2** *Exclusion Restriction*

$$Cov(z, \varepsilon_{ij}) = 0. \tag{3}$$

Like all other exogenous regressors, $z$ also has to be uncorrelated with the error term in the structural equation (1). The determination of the instrument outside the model (exogeneity) reflects the need for a random assignment of the instrument and the fact that the instrument affects the outcome variable, $y_{ij}$, only through the first-stage channel (Angrist and Pischke, 2009, 117).

If both assumptions are fulfilled IV estimation consistently estimates the parameter of interest $\theta$ (Wooldridge, 2010, 91). Only the first of the two assumptions can be empirically checked. This is done by estimating the following first-stage regression:

$$\left(\frac{H}{TP}\right)_j = \pi_0 + \pi_1 z_j + X_{ij}'\delta + v_{ij}. \tag{4}$$

Afterwards, one has to test the null hypothesis of $H_0 : \pi_1 = 0$. If this can be rejected at all significance levels than instrument relevance is given. The exogeneity assumption of the instrument cannot be checked and has to be reasoned by economic intuition. Exogenous instruments are often found in the context of random or natural experiments (Angrist and Krueger, 2001, 73). Wooldridge (2010, 94) states that "(...) a natural experiment occurs when some (often unintended) feature of the setup we are studying produces exogenous variation in an otherwise endogenous explanatory variable (...)". In order to assess the degree of introduced exogenous variation profound knowledge of the historical and institutional setting is necessary (Angrist and Krueger, 2001, 73).

Hornung (2014a, 100) argues that such an event can be found in the context of the Thirty Years' War and the Black Death. Due to Thirty Year's War, strong negative population shocks occurred from 1618 to 1648. Even though the war itself might have been endogenous, population losses during the time were partly driven by the Black Death. The author argues that the Black Death on the other hand is independent of economic conditions (Hornung, 2014a, 115). Instead of population or town size, the human-to-rat-ratio seems to determine the speed and magnitude of infection and population losses (Hornung, 2014a, 103).[3] The variable $PopLosses_j$ which quantifies the town population loss, therefore, is proposed as an instrumental variable candidate for $z_j$.
Using the first-stage fitted values, the second-stage IV regression equation becomes:

$$y_{ij} = \alpha + \beta_l l_{ij} + \beta_k k_{ij} + \beta_m m_{ij} + \theta\left(\frac{\hat{H}}{TP}\right)_j + X_j'\gamma + \varepsilon_{ij}. \tag{5}$$

Hornung constructs three different instrument measures. The first instrument is the raw unadjusted population loss data from Keyser (1939-1974) (qtd. in Hornung (2014a, 101)). Results from this specification are depicted in columns two and three of table 2. The instrument is significant at the 10 percent level and is positive partial correlated with the Huguenot population share. The instrumented regressor is highly significant and indicates that economic output grows on average by 3.5 percent when the French Protestant share increases by one percent. The second construction interpolates the data by using population growth rates taken from Pfister (2007) (qtd. in Hornung (2014c, 3)). This is done in order to use the exact time period from 1625 to 1652 for which Brandenburg was involved in the war (Hornung, 2014c, 3). (4) and (5) of table 2 show that the instrument is

[3] Hornung (2014a, 103) states that "(...) human mortality rates of Black Death epidemics depend not only on the density of humans but, more important, on the density of rats and rat fleas.".

more significant in the first stage and that the estimated effect in the second stage has decreased slightly. The last instrument variation is constructed by aggregating three sources from Keyser (1939-1974), Behre (1905) and Wohlfeil (1976) (qtd. in Hornung (2014a, 101)). This extends the sample by 36 observations, which is very important due to the high variance of IV estimators in small samples (see figure 3 part c). (6) and (7) of table 2 show that the estimator in the second stage is only half as large as in the two other variations. Furthermore, it is only significant at the 5 percent level. It remains unclear whether the aggregated data is more reliable than the single source by Keyser (1939-1974) (qtd. in Hornung (2014a, 101)). Changes in the coefficient give cause for concern whether the results are robust to the data. Comparing OLS and IV estimates reveals that unaggregated IV estimates are more than twice as large as OLS estimates. This points to negative selection bias in the OLS estimate. It may be possible that Huguenots did not "randomly assign" themselves to their settlement destination, but rather that less diffusing Huguenots were drawn to Prussia. IV "purges" this endogeneity and increases the estimated effect.

Previously presented results demonstrate that there is a positive relationship between the observed population share in towns and the economic output of firms in these towns. This indicates that the sharing of Huguenot knowledge increased productivity. Whether this relationship is causal depends on the reliability of the IV assumptions, the used data and the conducted inference. The following section digs deeper into these three topics.

# 3 Critical Examination and Extending Computations

The following chapter takes an in-depth look at three crucial pillars of the analysis of Hornung: The standard errors used to compute inference, the validity of the exclusion restriction, and missing data imputation methods. Building on theoretical results, I extend Hornung's analysis by computing and interpreting block bootstrap standard errors, weak instrument statistics, and regressions with different imputation methods.

## 3.1 Comparison of Cluster-Robust and Cluster-Bootstrap Standard Errors

Usual heteroscedasticity-robust standard errors may severely underestimate the true square-root of diagonal elements of the variance-covariance matrix (Angrist and Pischke, 2009, 293). This is due to the fact that real-life microeconometric data is rarely identically and independently distributed (iid). In his famous paper, Moulton (1986, 387) has shown that correlation of observations within groups/clusters and independence across clusters can lead to large distortions in the variance-covariance matrix estimator. This results in bad inference and wrong conclusions regarding the significance of the estimated parameter of interest. The following subchapter analyses two potential solutions to this so-called

"Moulton problem" (Angrist and Pischke, 2009, 294): Cluster-robust standard errors and cluster- or block bootstrap standard errors. I describe potential advantages and disadvantages of both algorithms. Afterwards, I provide and interpret the empirical results using the unique dataset from Hornung (2014b).

### 3.1.1 Cluster-Robust Standard Errors

Hornung (2014a, 103) identifies the potential problem of individual firm observations being correlated within towns and independent across towns. This may be due to similar economic or institutional settings that affect all manufactories inside but not outside of a specific town. Ignoring this dependence can have big implications for the significance of the estimated effect.

In a setting with clusters/towns $j = 1, 2, ..., J$ the square-root of the following ratio is called the general (with varying exogenous regressors $x$) Moulton factor (Angrist and Pischke, 2009, 311):

$$\frac{V(\hat{\theta})}{V_{clu}(\hat{\theta})} = 1 + \left[\frac{V(n_j)}{\bar{n}} + \bar{n} - 1\right] \rho_x \rho_\varepsilon, \tag{6}$$

where $V(\hat{\theta})$ denotes the true variance of the estimated parameters, $V_{clu}(\hat{\theta})$ is the OLS variance of the slope coefficient, $V(n_j)$ is the variance of cluster sizes and $\bar{n}$ denotes the average town sample size. Furthermore, $\rho_x$ and $\rho_\varepsilon$ denote the within-cluster correlation of all the regressors and the error, respectively.

One can directly see that the relative distortion between true variance with clustered observations and conventional OLS variance increases in the intraclass correlation of regressors and the error and in the number of observations in each cluster (given $n > 1$) (Cameron and Miller, 2015, 6). It is important to notice that the cluster problem vanishes when either $\rho_x = 0$, $\rho_\varepsilon = 0$ or $n_j = 1$ for all $j \in J$ (Angrist and Pischke, 2009, 311). Hence, the conducted inference is farther off the higher the dependence of observations within clusters and the larger the average cluster size. Keeping the overall sample size constant and increasing the average number of observations in each sample means that there are fewer clusters (Angrist and Pischke, 2009, 310). Therefore, the observations are even less independent (Angrist and Pischke, 2009, 310). In order to circumvent this problem one has to find the smallest cluster level for which interclass independence can be assured and cluster accordingly. Hornung (2014a, 103) argues that this level is the town level. Accounting for the dependence within towns one can instead estimate the following cluster-robust variance matrix (Cameron and Miller, 2015, 8):

$$\hat{V}_{clu}(\hat{\theta}) = (X'X)^{-1} \hat{B}_{clu} (X'X)^{-1} \text{ with } \hat{B}_{clu} = \sum_{j=1}^{J} X_j' \hat{\varepsilon}_j \hat{\varepsilon}_j' X_j, \tag{7}$$

where X denotes the design matrix including all regressors. Hornung provides for all of

his presented regression results cluster-robust standard errors.[4] Columns 1 and 2 of table 3 compare standard White-Huber heteroscedasticity robust standard errors with cluster-robust standard errors for one extensive OLS baseline specification. The standard error drops from 0.5 to 0.23 when accounting for town dependence. The estimator is more precise when factoring in, that firms in the same town faced similar constraints.
I take into doubt whether this really ensures independence across interclass observations. I imagine that town borders during this post-war time were in a state of flux. Therefore, it may be better to find a group identifier that is more time-invariant. I suggest to cluster at a higher level, e.g. the county or regional level.

### 3.1.2 Cluster-Bootstrap Standard Errors

By conducting Monte Carlo simulations, bootstrapping approximates the distribution of a statistic (Cameron and Trivedi, 2005, 357). This can be with sampling from the observed data (nonparametric bootstrap with replacement) or a fitted distribution of the sample (Cameron and Trivedi, 2005, 357). The computational power needed in order to implement the simulations is becoming more and more possible with progress in information technologies (Cameron and Trivedi, 2005, 357).

Cameron, Gelbach and Miller (2008, 420 et sqq.) have shown that a special form of nonparametric bootstrap, namely block or cluster-bootstrap, usually performs better than standard cluster-robust standard errors.[5] Cluster bootstrap is a variant of the bootstrap algorithm in which the correlation structure within classes is kept in place (Bertrand, Duflo and Mullainathan, 2004, 265). This is done by sampling observations that belong to the same cluster together in the same "block" (Bertrand, Duflo and Mullainathan, 2004, 265).

The following equation shows how to compute the cluster bootstrap estimate of the variance-covariance matrix (Cameron and Miller, 2015, 12)

$$\hat{V}_{clu-boots}(\hat{\theta}) = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_b - \bar{\hat{\theta}})(\hat{\theta}_b - \bar{\hat{\theta}})', \tag{8}$$

where $B$ is the number of bootstrap iterations, $\hat{\theta}_b$ represents the estimator-vector in the b-th replication and $\bar{\hat{\theta}}$ is the average of all $B$ estimators.
Like standard methods (heteroscedasticity-robust or cluster-robust), bootstrapping also relies on large sample properties (asymptotics) and is only approximate in finite samples (Cameron and Trivedi, 2005, 357). Furthermore, a problem that remains with a small number of clusters is that the true size of the test conducted with bootstrapped standard errors will differ from the used significance level (Cameron et al., 2008, 417).

[4]This is easily done by adding the option `cluster(groupid)` in Stata after the specific regression command where groupid specifies the unique town identifier.

[5]For detailed information on the bootstrap algorithm the interested reader is referred to Cameron and Trivedi (2005, 360), Efron and Tibshirani (1994, 47) or Appendix A.

Concerning Hornung's analysis, cluster-bootstrap standard errors are almost seven times larger than normal cluster-robust standard errors (see table 3). Regardless of replication size or seed the coefficient is insignificant at every common level. This result cannot be reproduced when using the imputed measure for ln input 1802 instead of the unadjusted (see table 1 column 6). Although Hornung included a dummy for imputed values (in order to make sure that imputed values do not drive the magnitude of the estimator), the precision of the estimator seems to be driven by imputation.

According to Efron and Tibshirani (1994, 52) 200 bootstrap replications are more than enough to ensure that asymptotics kick in. Cameron and Miller (2015, 12), on the other hand, recommend using 400 samples for results that shall be published. Columns 3 to 6 of table 3 show different cluster-bootstrap algorithm specifications. One can see that estimates do not vary significantly when increasing the iteration number.[6] Changing the seed (for random number generation), on the other hand, introduced slight variation.

All in all, the analysis of different standard errors weakens Hornung's results. All OLS specifications without imputed input values produce insignificant estimates of the parameter of interest (see table 1). This problem does not apply when substituting unadjusted input values with multiple imputed ones. Furthermore, IV regressions with cluster-bootstrap standard errors (see table 2) produce estimates that are insignificant at every common level and instrument relevance is violated in the first-stage regressions. Although one can control for the magnitude of the estimator, imputation might still drive inference.

## 3.2 Analysis of Weak Instrument Statistics

Ever since the ground-breaking work by Bound, Jaeger and Baker (1995) microeconometricians have to be aware of the consequences of so-called weak instruments.

When an instrumental variable candidate is valid, IV regressions produce consistent estimators (Cameron and Trivedi, 2009, 194). Consistency describes the fact, that the estimators density collapses to a spike when the sample size tends to infinity. Part a) and c) of figure 3 depict two scenarios of consistent estimator simulations: OLS and IV with all assumptions being fulfilled. With increasing sample size the variance of both estimators shrinks and the density collapses at the true population parameter.

It is well known that even though IV produces consistent estimates, they might still be biased in small samples (Bound et al., 1995, 449; Cameron and Trivedi, 2009, 194). When there is a weak instrument (poor first stage regression fit) this bias might even be larger than the OLS bias. In fact Bound et al. (1995, 444) have shown that with a single instrument the relative inconsistency of IV to OLS can be written as

[6] Cluster bootstrap standard errors can be computed using the Stata option `vce(bootstrap,rep(400) seed(100)) cluster(groupid)`. In order to fully replicate the findings one should specify the seed.

$$\frac{plim\ \theta_{IV} - \theta}{plim\ \theta_{OLS} - \theta} = \frac{\rho_{z,\varepsilon}/\rho_{x,\varepsilon}}{\rho_{x,z}}. \tag{9}$$

The ratio increases in the endogeneity of the instrument ($\rho_{z,\varepsilon} > 0$). It decreases in the endogeneity of the regressor $x$ and in the correlation of the instrument with the endogenous regressor. A bad first stage fit (small $\rho_{x,z}$) implies that the relative inconsistency is very sensitive to even small violation of the exclusion restriction (which implies $\rho_{z,\varepsilon} \neq 0$) (Bound et al., 1995, 444). Such a situation is depicted in part d) of figure 3.

The reduced form regression reported in table 2 column 1 estimates a positive partial correlation between the aggregated measure of the instrument and the dependent variable from the structural equation which is significant at the 10 percent level. Hornung (2014a, 114) states that the exclusion restriction might be violated if towns were systematically differently affected by the population loss. This would imply a non-random assignment of the instrument. Hornung (2014a, 115) uses economic reasoning and several robustness checks in order to establish a non-endogenous occurrence of aggregated population loss. He finds that population loss is uncorrelated with many variables that indicate economic conditions which could have also driven or dampened the spreading of the Black Death. It is unclear whether or not these tests are sufficient enough to ensure instrument exogeneity. Town characteristics such as the number of physicians, hospitals or academic institutions remain unobserved due to data limitations. It is obvious that these aspects negatively influenced the number of deaths. However, if hospitals also positively affected the economic output of firms in these towns, then the estimated coefficient would be downwards biased when using IV:[7]

$$Corr(PopLoss_j, Hospitals_j) < 0 \rightarrow Corr((H/TP)_j, Hospitals_j) < 0 \tag{10}$$

$$Corr(y_{ij}, Hospitals_j) > 0 \Rightarrow E(\hat{\theta}) - \theta < 0. \tag{11}$$

Many rules of thumb quantitatively define a weak instrument. Table 4 reports some common measures. Shea (1997, 349) suggested the partial R-squared as an informal test for weak instruments, which does not control for type I error, based on an empirical statistic. Following Kennedy (2013, 49) it can be be computed in this way:

$$\text{partial } R^2 = \frac{R^2_{unr} - R^2_{res}}{1 - R^2_{res}}, \tag{12}$$

where $R^2_{res}$ is the $R^2$ from a first stage regression in which we abstract from the instrument and only use the other regressors to explain the instrumented variable and $R^2_{unr}$ is the unrestricted $R^2$ from a first stage regression of all exogenous regressors and the instrument. The partial $R^2$ describes the explanatory power of the instrument after having excluded

[7] Here I assume that $Hospitals_j$, a measure of the number of Hospitals in town $j$, only affects the Huguenot population share through the instrument, $PopLoss_j$. Otherwise the direction of the bias would be ambiguous.

the influence of other regressors (Cameron and Trivedi, 2009, 196).
Table 4 shows that the partial $R^2$ in Hornung's case lies between 0.15 and 0.3 and increases when interpolating and aggregating the instrument data. It is hard to judge whether this is large enough because there is no clear-cut threshold in the literature (Cameron and Trivedi, 2009, 196). Such a small partial $R^2$ bears the risk that even when the instrument suffers from small endogeneity, the relative inconsistency between IV and OLS is going to be large.
Staiger and Stock (1997, 557), on the other hand, proposed to declare an instrument to be weak if the first-stage F-statistic is smaller than 10. They derived that the relative bias of IV to OLS is proportional to the inverse of the first stage F-statistic testing the joint significance of excluded instruments:

$$\frac{E(\hat{\theta}_{IV})}{E(\hat{\theta}_{OLS})} - 1 \approx \frac{1}{F}. \tag{13}$$

One can directly see that strong instruments require large first stage F-statistics. The Kleinbergen-Paap F-statistics (displayed in table 4) surpass the threshold of 10 only when changing the instrumented variable. Instead of instrumenting the population share of Huguenots in 1700, Hornung (2014a, 109-110) has to use the natural logarithm of Huguenots in the textile industry in 1700 as the instrumented regressor. This changed regressor measures French Protestants contributions to the production process more directly than before. Therefore, it is unclear whether Hornung really estimates the effect of first-generation technology diffusion. Huguenots who worked in the textile industry did not only share their knowledge, but also worked and produced goods. The channel of affection would broaden and the estimated model is likely to answer a different question: How did Huguenot immigrants *and their work in the textile industry* affect Prussian production?

## 3.3 Alternative Methods for Imputation of Missing Data

The simplest but also most dangerous way of dealing with missing data is to drop all observations that do not contain information for all included regressors (complete-case analysis) (Gelman and Hill, 2006, 531). This is a luxury that can only come up in big data sets. Hornung (2014b) created a unique data set that contains at a maximum 750 observations. The smallest sample analyzed in IV regressions consists of only 150 observations. As I have pointed out, instrumental variables estimators, although unbiased, have a large variance in small samples (see figure 3). Therefore, dropping observations is even more dangerous for the identification strategy than in an usual OLS setting.

Hornung (2014a, 103) identifies the missing-data problem for 15 percent of the observations for the input measure. He makes use of univariate multiple imputation (Hornung, 2014c, 2).[8] It is crucial to consider the underlying mechanism that causes the missingness because the assumptions derived are important for the method conducted to fill in the gaps. In general, one has to differentiate between two missing-data mechanisms: Missingness completely at random (MCAR) and missingness at random (MAR) (Rubin and Little, 2002, 12). Following Rubin and Little (2002, 12) one can characterize data as MCAR if the following conditional density $f(.)$ holds for the data at hand:

$$f(M|R,\phi) = f(M|\phi) \text{ for all } R,\phi, \tag{14}$$

where $M$ denotes the missing-data indicator matrix and R is a matrix including the complete data set with $K$ variables. Its $ik$-th element is 1 if the value of $r_{ik}$ is missing and 0 otherwise (Rubin and Little, 2002, 4). $\phi$ are unknown parameters. The equation states that the conditional distribution of the missing-data indicator matrix has to be independent of missing or observed values $R$ (Rubin and Little, 2002, 12). Therefore, the probability of missingness is equal for all observations in the sample (Gelman and Hill, 2006, 530). If the missing-data mechanism is completely at random, the observations with missing values can simply be dropped from the sample without biasing the analysis (Gelman and Hill, 2006, 530).

Usually MCAR is a too strong assumption. Missingness of the value of input materials could for example depend upon the actual value. Firms could have been afraid that competitors observed the survey information and, therefore, either provided wrong or no answer at all. In this case MCAR would be validated and the missingness would depend on the missing value itself (Gelman and Hill, 2006, 530).

Missingness at random can be formally stated like this (Rubin and Little, 2002, 12):

$$f(M|R,\phi) = f(M|R_{obs},\phi) \text{ for all } R_{mis},\phi. \tag{15}$$

The conditional distribution of the missing-data indicator matrix does not depend upon the missing values of $R$ but depends on the observed values, $R_{obs}$ (Rubin and Little, 2002, 12). If one would drop the missing observations in such a case sample-selection bias would be introduced. This is especially a problem if the remaining observations do not have representative values of the regressor conditional on the other regressors (Cameron and Trivedi, 2009, 47). Many imputation methods such as the ones exploited by Hornung make use of this less restrictive assumption.

I am interested in the robustness of Hornung's results to different imputation methods. As previously stated, inference seems to be driven by imputed values of input material. Cluster-bootstrap standard errors reveal that OLS estimation with non-imputed values are not significant, whereas the ones with multiple imputed values remain significant at the

[8]Unfortunately Hornung does not provide the replication code for the imputation itself. I am unable to replicate the exact values of Hornungs imputed input/material measure. Attempts following Hornung (2014c, 2) can be found in the replication material.

one percent level (see table 1).

The simplest algorithm for imputing values is mean imputation. In this case, all missing values are replaced with the sample average of the imputed variable (Gelman and Hill, 2006, 532). Part c of figure 4 depicts the frequency distribution (or histogram) of the mean imputed variable. One negative feature of mean imputation is that the variance is drastically reduced (Gelman and Hill, 2006, 532-533). In this case, the sample variance decreases almost by one-third. Furthermore correlations will be reduced which might make this method bad for our regression application (Gelman and Hill, 2006, 533).

The next algorithm I used is simple random imputation. In these cases, Gelman and Hill (2006, 534) state that the missing value is changed to the value of a randomly sampled non-missing observation in the sample. The histogram depicted (part d of figure 4) looks very similar to the multiple imputed data from Hornung (part b of figure 4) and the mean and second central moment are very similar. Still regression results come up with very different estimates of coefficients and standard errors (column 3 of table 5) which indicate an insignificant effect.

One can also make use of the other variables contained in the data set and compute different predictive regression models (Gelman and Hill, 2006, 535-536) for the missing parts of the variable. In a baseline regression I compute the following model

$$m_{ij}^{mis} = \alpha + \beta_y y_{ij} + \beta_k k_{ij} + \beta_l l_{ij} + X_{ij}'\delta + u_{ij}, \tag{16}$$

where $X_{ij}$ contains multiple different material specification dummies.[9] Afterwards, I compute fitted values for the missing values and impute them. It is important to remember that the goal of this specification is not to come up with an estimator with beautiful asymptotics, but to accurately predict the missing values. Gelman and Hill (2006, 537) state that it is, therefore, possible to include more regressors than usual when using this method.

Following Gelman and Hill (2006, 536), it is also possible to introduce ambiguity to this imputation by adding a prediction error to the equation. Part e of figure 4 displays the corresponding histogram and columns 4 and 6 of table 5 the OLS regressions for these input value variations. Only the random regression imputation produces an estimate that is significant at the 10 percent level (when using cluster-robust standard errors).

The last imputation algorithm I apply is iterative regression imputation. This method makes use of two variables with missing data and iteratively uses one of the variables to predict the values of the other variable (Gelman and Hill, 2006, 539). This is done by using the fitted values of one regression model in the model for the other missing-data variable. In general, this process continues until the values change minimally (Gelman and Hill, 2006, 539). The frequency distribution of this approximately converged value

[9] Dummies include indicators for wool, linen, cotton, silk, hats, socks, "tuch" production and the percentage of merino sheep, 1816 (county).

(using the aggregated population loss measure as the second variable) is displayed in part f of figure 4. Also this regression predicts an insignificant relationship between the presence of Huguenots and the economic output of firms (column 5 of table 5).

All in all, I find that Hornung's results are very sensitive to different imputation methods. Because Hornung does not provide the exact algorithm with which he came up with imputed values,[10] it remains unclear how robust his results really are. I find that there is only a positive relation between the population share of Huguenots and manufactory output if one imputes the input value with a regression or random regression imputation technique and uses cluster-robust standard errors for inference. All other methods (and all inference based on block bootstrap standard errors) yield insignificant estimates of the parameter of interest. Furthermore, the dummy included to control for the effects of imputation has a significant influence in almost all cases. It is also important to remember that standard errors of coefficients estimated with imputed variables are generally downwards biased (Gelman and Hill, 2006, 532). This is due to the fact that the massive uncertainty introduced by imputation of 15 percent of one regressor is neglected in the standard error estimation. This remains unmentioned by Hornung.

All results presented in this chapter provide contradictory evidence regarding the conclusions of Hornung. Therefore, the next chapter consults other literature developments that might help to enlighten the relationship of interest.

# 4 Literature Analysis, New Considerations, And Open Questions

After having analyzed the strengths, weaknesses, and implications of Hornung's analysis, I now turn to the relevant literature which deals with the implications of high-skill immigration. First, I describe important contributions and their context. In order to assess the validity of Hornung's results for our modern world, I outline the different channels of knowledge transfers that determine technology diffusion nowadays. Furthermore, I highlight open research topics such as the speed of knowledge adjustments, social networks, and the importance language assimilation.

## 4.1 Comparative Analysis of the Relevant Literature

Most of the literature that deals with the implications of immigration focuses on specific economic effects such as labor supply shocks, wage adjustments for both natives and immigrants, and changes in patenting (see Kerr and Kerr (2001) for an overview). Similar

[10]Hornung (2014c, 2) states that he makes use of the univariate multiple imputation method and lists the explanatory variables used in a random regression. Hornung does not provide the exact number of imputations or the seed and my attempts to replicate the measure do not compute the same values.

to the branch that deals with topics in innovation, Hornung studies the effects of human capital externalities due to high-skill immigration. Instead of analyzing changes in innovation output measures such as patents, he focuses on adjustments in overall economic output concerning different manufacturing industries.

Moser et al. (2014, 3223) examine the quantitative effects of German-Jewish immigration to the United States on US invention after escaping from the Nazi dictatorship. IV regressions indicate that escaped Jewish chemists increased patenting by US inventors per USPTO (United States Patent and Trademark Office) class and year by 17 percent (Moser et al., 2014, 3240). The authors state that this increase is probably not caused by knowledge spillovers introduced by the immigrants (Moser et al., 2014, 3245). Instead they argue that a "peer effect" caused by the immigration of several prestigious chemists, attracted many US scientists into emigres' fields of study (Moser et al., 2014, 3246). Therefore, the immigrants did not directly increase the productivity of US scientists but instead shifted their focus of attention (Moser et al., 2014, 3253). This result dampens the idea of diffusion and instead highlights the importance of group effects. Applying this result to Hornung's analysis brings up the following question: Did Huguenots not only share knowledge, but also incentivize more productive natives to work in the textile industry? Not controlling for this effect would lead to an upwards bias in the estimate of the effect of technology diffusion.

An important related study by Borjas and Doran (2012, 4) analyzes the developments of US mathematicians' productivity after the immigration of Soviet colleagues to the US. Again, the group of immigrants was composed of highly talented and skilled academics (Borjas, 2014, 185). The authors found that Soviets suppressed native scientists from the job and publication market (Borjas and Doran, 2012, 39). US mathematicians who were directly confronted with research of the immigrants suffered a negative productivity shock, were less likely to publish first-class papers and often had to switch affiliated institutions (Borjas and Doran, 2012, 4-5). Borjas and Doran (2012, 31) state that the long-term effects are likely to be even worse due to the sensitive nature of academic careers. But it is important to notice that this study only deals with the very specific market of academia (Borjas, 2014, 191). Hence, Hornung extends the literature by studying the effects of knowledge transfers due to immigration on a broader scale: industrial production.

Borjas (2014, 172) also emphasizes that there may be severe differences between the short-run adjustments of high-skilled worker migration and the long-run effect. He describes the short-run developments as a "(...) race between the spillover effect and the law of diminishing returns (...)." (Borjas, 2014, 172). Due to the rigidness of the production process in the short-run, the marginal product of an additional unit of effective labor is shrinking. In the long-run substitution and capital adjustments kick in, which outweigh the resource scarcity (Borjas, 2014, 172).

In the long-run Moser et al. (2014, 3251) also find that co-inventors of German-Jewish chemists were more productive in their further career. This helped to form a new and young generation of scientists that increased patenting (Moser et al., 2014, 3253). It seems likely that this might have also happened in Prussia with the apprenticeships that already helped to decrease prejudices. But the competitive aspects highlighted above, indicate that the natives often times were not able to fully benefit from the knowledge of high-skilled migrants.

Waldinger (2010), on the other hand, studies the effects of a negative supply shock in high-skilled academics on the productivity of scientists in the country of origin. He examines the changes in productivity of doctoral students in mathematics after many well-known Jewish mathematicians had to leave Germany in 1933 after Hitler came to power. Waldinger (2010, 812) finds that there exist strong positive human capital externalities. Students, who were part of departments with severe losses, experienced significant decreases in their productivity. Borjas (2014, 183) concludes that this study reveals a negative effect on students of high-skill workers but not on colleagues, which where locally near. Therefore, the quality of training is very important even in highly skilled job markets such as academia (Waldinger, 2010, 791). Again, the importance of mentorship has to be highlighted (Borjas, 2014, 181).

However, Lausberg (2007, 209) reports discrimination of Huguenots in Prussia. Moser et al. (2014, 3223) state that Jewish immigrants to the US faced anti-Semitism. Despite this, they still estimate a strong effect of immigrants on innovation output. It is hard to control for discrimination in such empirical settings. Therefore, the regressors of interest are likely to suffer from OVB. A measure of the intensity of intolerance should be negatively correlated with regressand and regressor in both cases. Therefore, the estimated effect is expected to be downwards biased.

In my opinion, one of the critical questions regarding the analysis of Hornung is what the influence of the population share of immigrants in the structural equation really measures. It is possible that Huguenots motivated Prussian manufacturers to work harder. Maybe Prussians were more productive because they feared the loss of their jobs. One thing is for sure: There may be other channels except for technology diffusion through which Huguenots might have affected the output of manufactories.[11] If an unobserved motivation measure is positively related not only to the firm-level output measure, but also to the population share of immigrants than the estimator is going to be upwards biased. It is likely that other unobservables introduce complex OVB, which cannot be fully solved by IV estimation. In my opinion, the partial correlation captured in the partial R-squared (analyzed in chapter 3.2) is not large enough to combat all of the indicated endogeneity.

The studies consulted above indicate that there are multiple channels through which im-

[11] Due to the nature of the data (cross-section) it is impossible to account for unobservable factors such as ability or motivation. Adding a time dimension to the data set could solve this problem by making use of fixed effects (Wooldridge, 2010, 282 et sqq.).

migration affects the transfer and creation of knowledge. The majority of the reviewed literature comes up with a small positive effect of immigration on the studied output. This may be due to two contrary effects: positive human capital externalities and a decreasing marginal product of the production factor labor. In my opinion, however, this is only half of the picture: It is crucial to study the long-run in which prejudices have dismantled and capital adjustments happened. In this context I highlighted the importance of mentorship and "peer effects". Furthermore, one has to look at both countries, the host country and the country of origin in order to assess the overall implications. It is hard to conclude with a general effect for the modern world due to the fact that most studies concentrate on specific historical events that led to the immigration of highly selective groups. Thus, in my opinion it is necessary to analyze differences between channels of knowledge transfer in the 17th and 21st century.

## 4.2 Consequences of Shifting Knowledge Diffusion Mechanisms

Hornung (2014a, 85) argues that the historical setting of Huguenot migration to Prussia is exceptionally suitable for studying the process of technology diffusion through immigration because it excludes modern channels of technology transfers such as the changing nature of mobility and modern means of communication. Transnationalism, globalization, the internet, social media and MOOCs (Massive Open Online Courses) changed the way every individual perceives and accumulates knowledge. Nonetheless, following Williams and Baláž (2008, 2), I argue that immigration and human mobility still are major channels through which knowledge diffuses.

In general, one has to differentiate between the transfer of tacit knowledge, which is individual and context bound and the transfer of explicit knowledge, which can be transmitted via formal education (Williams and Baláž, 2008, 40). Furthermore, it is possible to sub-divide tacit knowledge into 4 classes: embrained, embodied, embedded and encultured knowledge (Williams and Baláž, 2008, 40-41). The embrained and embodied parts of tacit knowledge are clearly inseparable from the human being. Hence, it can only be fully diffused by corporeal mobility (Williams and Baláž, 2008, 43). Williams and Baláž (2008, 45) define immigration partly as the transfer of a special combination of embrained, embodied, encultured and embedded knowledge. Furthermore, they state that knowledge is always relational (Williams and Baláž, 2008, 60). By sharing it the potential economic value is inevitably transformed, too. Essential for the economic outcome of this transaction is the recombination of the knowledge components (Williams and Baláž, 2008, 45). New forms of communication enhanced the spreading of codified knowledge. But there still remain parts of tacit knowledge that can only be shared by face-to-face contact (Williams and Baláž, 2008, 65). One special channel threw which ideas diffuse are so-called knowledge brokers (Williams and Baláž, 2008, 75-76). Knowledge brokers are highly intelligent and socially capable individuals who are able to translate, coordinate and align between different perspectives of thought (Williams and Baláž, 2008, 77). Instead of generating knowledge they often convert ideas across boundaries and thereby span them (Williams and Baláž, 2008, 77).[12]

As certain elements of tacit knowledge can only be shared by experience or face-to-face communication and a process of recreation by knowledge brokers, I conclude that technology diffusion threw immigration is going to have a positive effect on aggregate economic output of the host country even in the face of new means of communication.

## 4.3 Speed of Adjustment, Social Networks and Language

Several aspects that Hornung did not examine are worth taking a closer look: In particular speed of adjustment in technology diffusion, the importance of social networks for integration and language assimilation play an important role.

Acemoglu (2009, 613) describes the process of diffusion as a logistic or sigmoid function: After a slow process of adoption a critical threshold of assimilation is reached. It is followed by a phase of rapid diffusion. In the end, the process slows down again. Furthermore, relatively underdeveloped countries should adjust at a relative faster pace (Acemoglu, 2009, 615). This would imply that the quantitative effect observed by Hornung is partly driven by the relative difference between Huguenot's and the Prussian knowledge. Again, adding a time dimension to the data would allow to study annual treatment coefficients and to analyze the dynamics of diffusion.

Furthermore, it would be very interesting to study the selective power of social networks. It might be possible that Huguenots chose to settle in towns in which there were already strong Protestant communities established. Portes (1995, 8) states that social networks are very important structures needed to understand economic transactions. On the one hand, networks give immigrants fast access to crucial adjustment information. On the other hand, they tend to decrease incentives for migrants to assimilate into the native's society because of perceived social stability.

Williams and Baláž (2008, 40) argue that the key to understanding the transmission of technology is the relation of knowledge and learning through language. There seems to be a short way from knowledge translation to actual creation (Williams and Baláž, 2008, 40). Viewing knowledge transfer as a process of translation creates questions about the importance of language concerning technology diffusion. Incentives that drive immigrants' learning of the native language are crucial for understanding not only the process of integration but also of face-to-face communication. Learning a language in general is a function of the individual's age, the "linguistic distance" between host and the immigrant's original language, the intention of staying and the degree of integration (Williams

[12] Popular examples include Matteo Ricci, Marco Polo, Enrico Fermi and Albert Einstein (Williams and Baláž, 2008, 78-79).

and Baláž, 2008, 28). There are two problems that may arise in this context. Immigrants who intend to return to their country of origin might have less intention to fully master the new language (Williams and Baláž, 2008, 29). Furthermore, the language learning function might experience a "levelling-off" effect (Williams and Baláž, 2008, 29): Instead of further advancing their language skills immigrants might settle at a satisfactory level rather than at their zenith. This effect could be even stronger when social networks are strong and there is less need to assimilate.

Providing incentives for immigrants to fully develop their potential, therefore, is an important policy target that needs to be further examined.

# 5 Conclusion

This thesis has theoretically and empirically analyzed the effects of high-skill immigration on the diffusion of new technologies. At the center of this analysis was the empirical analysis conducted by Erik Hornung. Examining Hornung's historical setting, I found that not only the exogenous orchestration of the Great Elector but also incentives such as trade connections and academic institutions influenced the settlement decision of Huguenots. Furthermore, I highlighted that Huguenots did not only receive benefits, but also were discriminated. Only in the long-run apprenticeships were able to culturally assimilate the immigrants. After giving an overview of the used data, the identification strategies employed by Hornung were motivated and established. I found that adding a control for workers in the regression equation severely decreases the estimator. This introduced questions regarding the elements of diffusion captured in the regressor. Comparing IV and OLS estimator showed that there is negative selection bias affecting the estimator of the parameter of interest. In order to assess Hornung's analysis it was necessary to study the two main underlying assumptions of instrumental variables estimation, namely instrument relevance and the exclusion restriction. Examining weak instrument statistics I found that the instrument is not capturing enough exogenous variation in the regressor to combat all potential threads of endogeneity. Furthermore, it remains unclear whether the robustness checks conducted by Hornung are enough to ensure the exogeneity of the instrument. Conducting inference with cluster- or block bootstrapped standard errors further weakened Hornung's results. Inference is driven by the imputation method employed by the author. In order to further check for robustness I used several different imputation algorithms and found that it is impossible to recreate Hornung's findings when deviating from his univariate multiple imputed measure of input values. Corresponding literature indicated that the intuition derived by Hornung may still be valid in the long-run. After several periods of capital adjustments, human capital externalities caused by outstanding migrants outweigh the negative effect of scarce resources. Furthermore, group-, mentoring and competitive effects are crucial. Even in our modern setting there are still parts of tacit knowledge that can be only shared via face-to-face interaction. Some interesting open research questions include the effects of social networks and knowledge brokers. Understanding how these two channels of diffusion work, is crucial to the success of not only technology adoption but also integration of immigrants. I conclude that Hornung's analysis suffers from severe statistical uncertainty. The alleged strength of the historical setting in circumventing several modern channels of diffusion also seems to be the biggest weakness of the empirical strategy. The data situation introduces to much ambiguity. The story told by Hornung seems to be partly supported by the literature and surely reflects political correctness but cannot be strengthened by the data.

## Tables and Figures

Figure 1: Scatterplot - Textile Industry

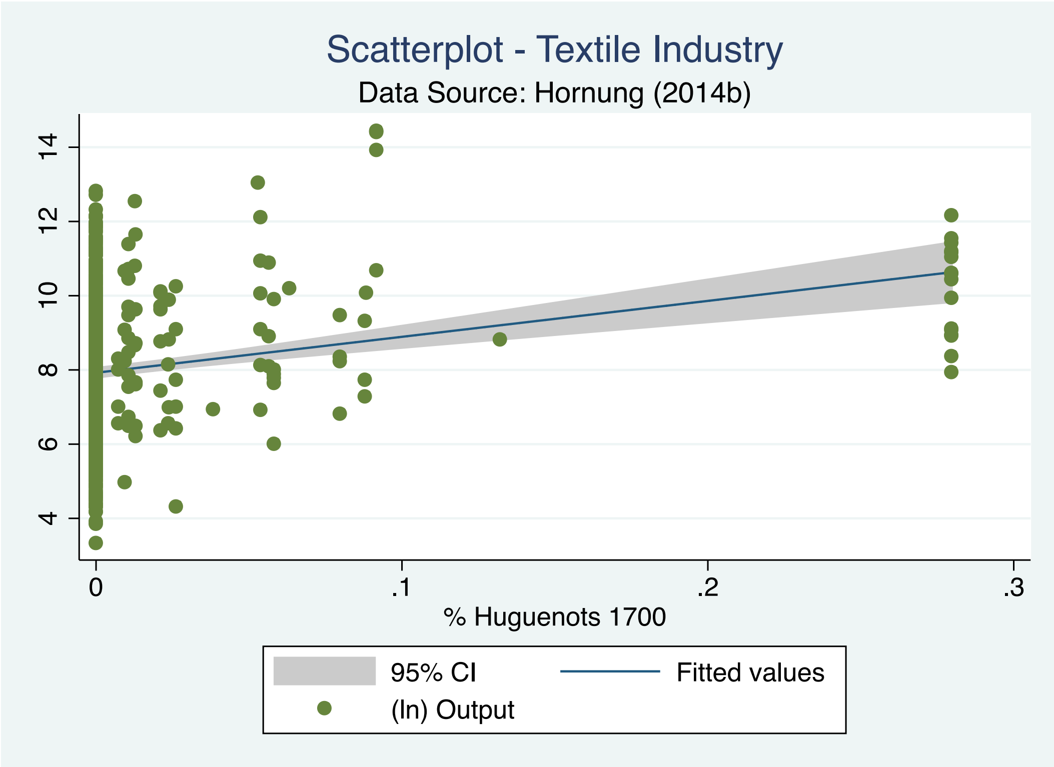



Figure 2: Scatterplot - Non-textile Industries

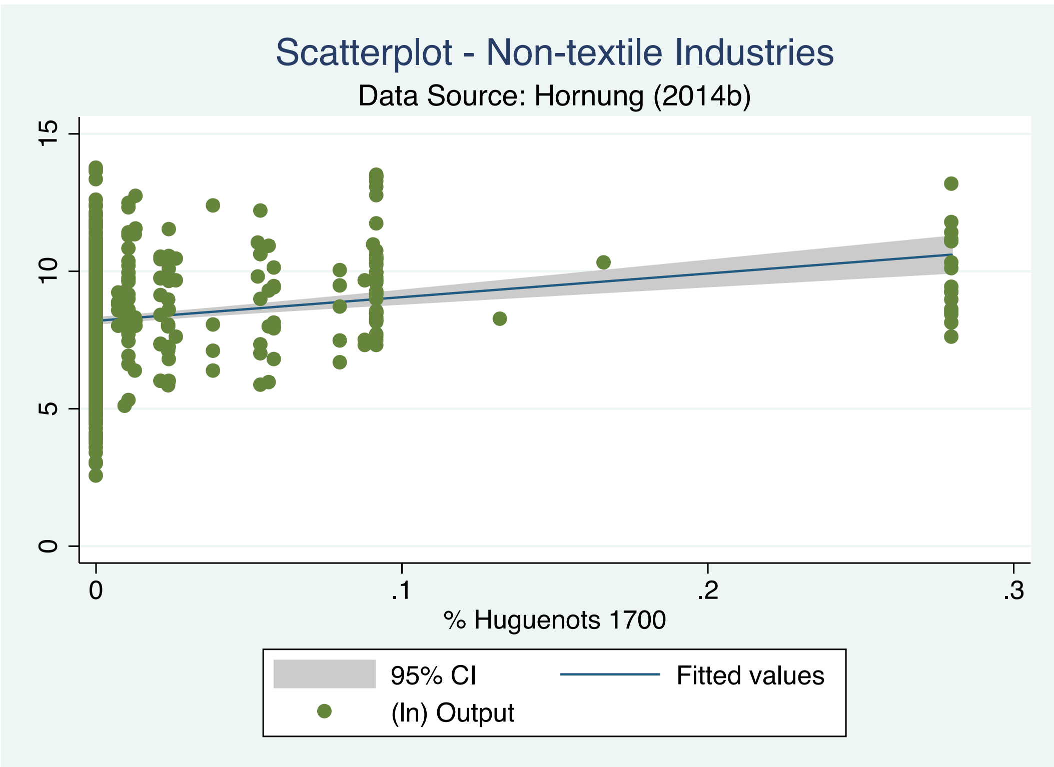
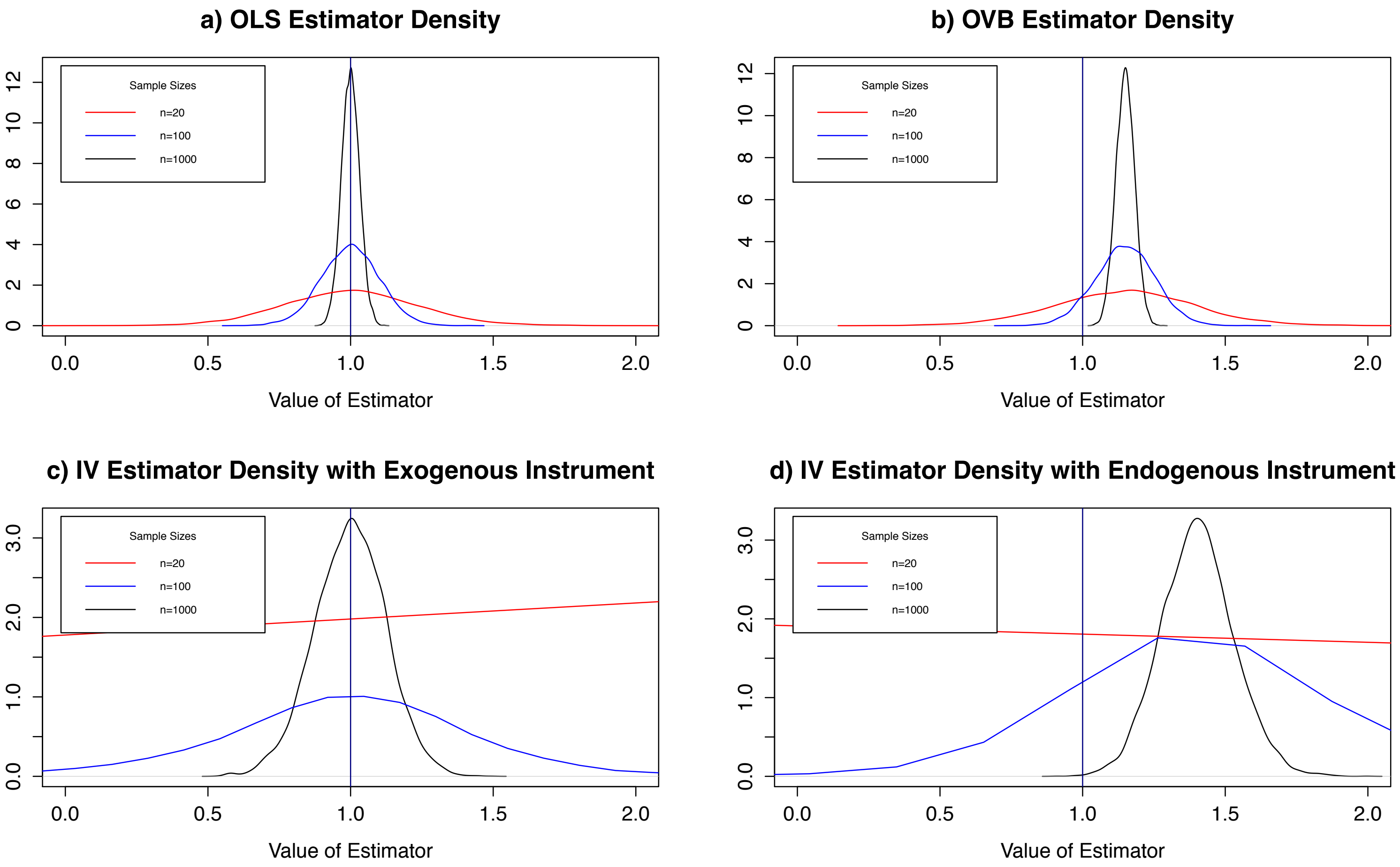
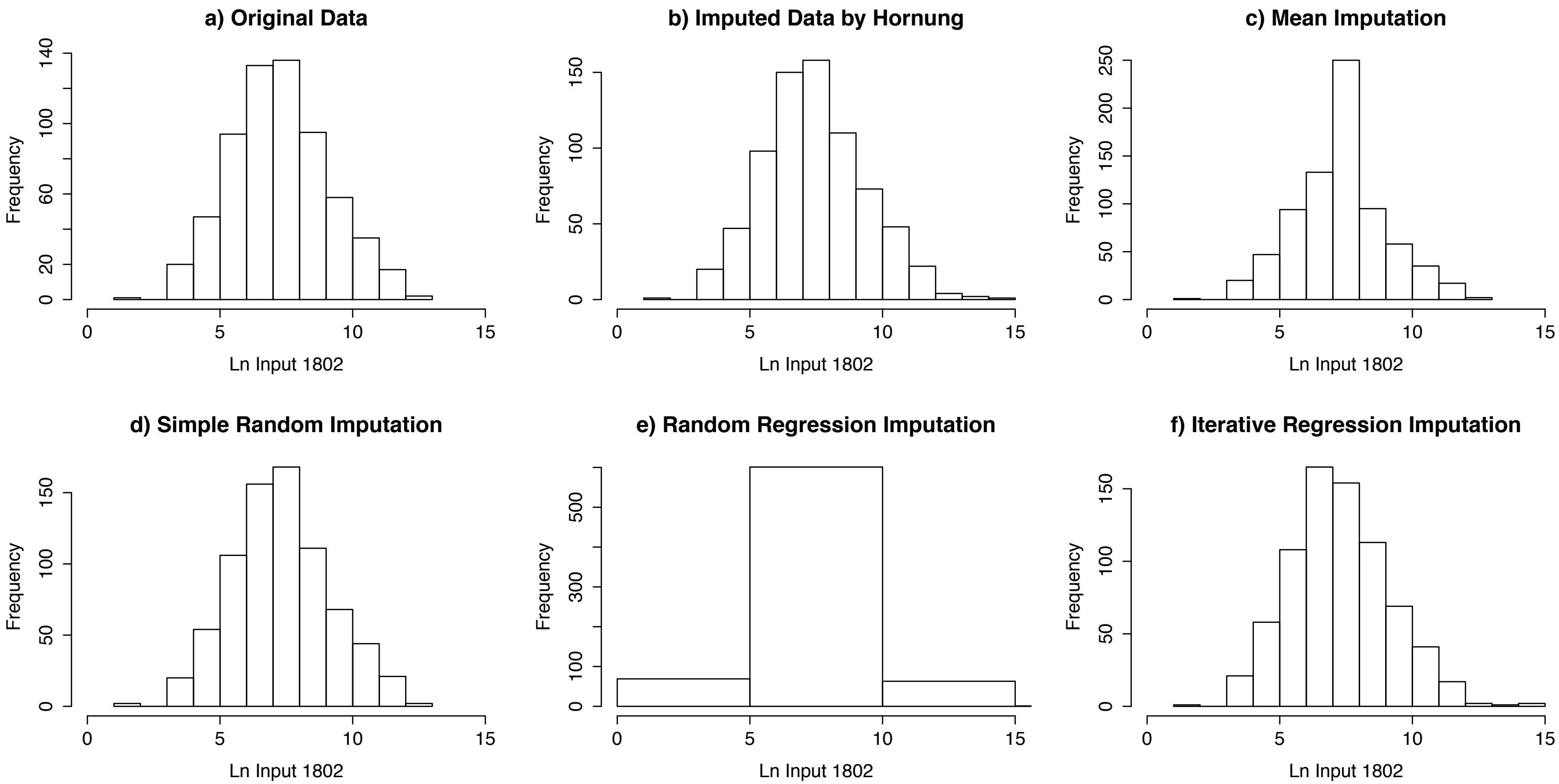


Figure 3: Comparison of Different Estimator Densities




*Notes:* Source: Own calculations. Monte Carlo simulation with 10,000 replications each. The true population parameter value is 1. Endogeneity in the instrument is introduced by correlating the instrument with a omitted variable (see Angrist and Krueger (2001, 79)). The correlation between the instrument and the omitted variable is 0.4. The exact replication code can be found in the supplementary material provided with this thesis.

Figure 4: Comparison of Different Imputation Methods




*Notes:* Source: Hornung (2014b) and own calculations following Gelman and Hill (2006, 529-542). Imputation methods for regressor ln input 1802: a) Original data from Hornung (2014b) b) Multiple imputed data from Hornung (2014b) c) Mean imputed data d) Simple randomly imputed data e) Random regression imputated data f) Iterative regression imputation. Regression imputations are performed using ln input 1802 as the dependent variable and computing an OLS model using the production indicators (wool, linen, cotton, silk, hats, socks, "tuch"), output 1802, the looms 1802 dummy, the number of workers 1802 and p.c. Merino sheep as predictors. The code can be found in the supplementary material provided with this thesis.

Table 1: Simple Ordinary Least Squares Regressions

| Dep. Variable: ln output 1802 | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Percent Huguenots | 9.6809 | 3.1814 | 1.4003 | 1.5071 | 1.4936 | 1.4588 |
| *Cluster-Robust* | (1.9641)*** | (0.7044)*** | (0.1723)*** | (0.1910)*** | (0.2324)*** | (0.2100)*** |
| *Cluster-Bootstrap* | (8.4687) | (3.019) | (1.2764) | (1.6764) | (1.6894) | (0.5397)*** |
| ln workers 1802 | | 0.9318 | 0.1255 | 0.1252 | 0.1251 | 0.1385 |
| *Cluster-Robust* | | (0.0227)*** | (0.0246)*** | (0.0244)*** | (0.0251)*** | (0.0222)*** |
| *Cluster-Bootstrap* | | ( 0.0224)*** | (0.0236)*** | (0.0236)*** | (0.0241)*** | (0.0228)*** |
| ln looms 1802 | | | 0.0406 | 0.0376 | 0.0377 | 0.0333 |
| *Cluster-Robust* | | | (0.0229)* | (0.0229) | (0.0233) | (0.0201)* |
| *Cluster-Bootstrap* | | | (0.0230)* | (0.0228)* | (0.0230) | (0.0204) |
| ln input/materials 1802 | | | 0.8066 | 0.8082 | 0.8082 | |
| *Cluster-Robust* | | | (0.0244)*** | (0.0244)*** | (0.0244)*** | |
| *Cluster-Bootstrap* | | | (0.0242)*** | (0.0240)*** | (0.0241)*** | |
| Not using looms (dummy) | | | 0.1367 | 0.1325 | 0.1326 | 0.1373 |
| *Cluster-Robust* | | | (0.0643)** | (0.0647)** | (0.0648)** | (0.0602)** |
| *Cluster-Bootstrap* | | | (0.0632)** | (0.0634)** | (0.0636)** | (0.0622)** |
| ln input/materials 1802 (imputed) | | | | | | 0.8054 |
| *Cluster-Robust* | | | | | | (0.0219)*** |
| *Cluster-Bootstrap* | | | | | | (0.0214)*** |
| Dummy for imputed values | | | | | | 0.0323 |
| *Cluster-Robust* | | | | | | (0.0342) |
| *Cluster-Bootstrap* | | | | | | (0.0357) |
| Intercept | Yes | Yes | Yes | Yes | Yes | Yes |
| Town and Availability Controls | No | No | Yes | Yes | Yes | Yes |
| Religion and Annexation Controls | No | No | No | Yes | Yes | Yes |
| Textile Production Control | No | No | No | No | Yes | Yes |
| Observations | 750 | 637 | 597 | 597 | 597 | 693 |
| Number of Clusters/Towns | 342 | 278 | 250 | 250 | 250 | 302 |
| $R^2$ | 0.0530 | 0.9621 | 0.9627 | 0.9630 | 0.9630 | 0.9669 |
| Adjusted $R^2$ | 0.0517 | 0.9618 | 0.9623 | 0.9625 | 0.9624 | 0.9664 |

*Notes:* Textile industry data is taken from Hornung (2014b). (1), (3), (4), (5), (6) are the same as in Hornung's (2014, 104) Table 2. Cluster-robust standard errors are clustered on town level. Cluster-bootstrap standard errors are computed using 400 replications. Town and availability controls consist of ln town population 1802 and Merino sheep, p.c. 1816. Religion and Annexation Controls consist of Percent Protestant 1816 and a dummy indicating whether the town did not belong to Prussia in 1700. Textile Production Control consists of a dummy indicating whether the town already had a relevant textile production in 1680. Standard errors are reported in parantheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Reduced Form and Instrumental Variables Regressions

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Reduced Form | First Stage | Second Stage | First Stage | Second Stage | First Stage | Second Stage |
| Dependent Variable: | ln output 1802 | PC H 1700 | ln output 1802 | PC H 1700 | ln output 1802 | PC H 1700 | ln output 1802 |
| Instrument Construction : | | Unadjusted | | Interpolated | | Aggregated | |
| Percent Huguenots 1700 | | | 3.4752 | | 3.3802 | | 1.6708 |
| *Cluster-Robust* | | | (1.1562)*** | | (1.1367)*** | | (0.8511)** |
| *Cluster Bootstrap* | | | (57.5474) | | (17.7894) | | (9.1279) |
| PC PopLoss 30 Years' War, aggregated | 0.1688 | | | | | 0.1010 | |
| *Cluster-Robust* | (0.0990)* | | | | | (0.0422)** | |
| *Cluster Bootstrap* | (0.1389) | | | | | (0.0572)* | |
| PC PopLoss 30 Years' War, Keyser unadjusted | | 0.0717 | | | | | |
| *Cluster-Robust* | | (0.0374)* | | | | | |
| *Cluster Bootstrap* | | (0.0511) | | | | | |
| PC PopLoss 30 Years' War, Keyser interpolated | | | | 0.0854 | | | |
| *Cluster-Robust* | | | | (0.0390)** | | | |
| *Cluster Bootstrap* | | | | (0.0516)* | | | |
| Intercept | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Additional Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 186 | 150 | 150 | 150 | 150 | 186 | 186 |
| Number of Clusters/Towns | 71 | 57 | 57 | 57 | 57 | 71 | 71 |
| $R^2$ | 0.9765 | 0.5931 | 0.9782 | 0.6005 | 0.9785 | 0.6108 | 0.9785 |
| Adjusted $R^2$ | 0.9750 | 0.5606 | 0.9765 | 0.5686 | 0.9768 | 0.5862 | 0.9768 |

*Notes:* Textile industry data is taken from Hornung (2014b). (2), (3), (4) and (5) are the same as Hornung's (2014, 108) columns 2, 3, 4 and 5 in table 4. (6) and (7) are the same as Hornung's (2014, 110) columns 2 and 3. Cluster-robust standard errors are clustered on town level. Cluster-bootstrap standard errors are computed using 400 replications. Additional controls that are not displayed are: Ln workers 1802, ln looms 1802, ln input/materials 1802 (imputed), Not using looms (dummy), ln town population 1802, Merino sheep, p.c. 1816 (county), Percent Protestant 1816, Not Prussia in 1700 (dummy), Dummy for imputed values, Relevant textile production 1680. Standard errors are reported in parantheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## Table 3: Comparison of Different Variance-Covariance Estimators

| *Panel A* | (1) | (2) | (3) |
|---|---|---|---|
| | White-Huber | Cluster | Cluster Bootstrap |
| | Heterosced. Robust | Robust | 50 Replications |
| Percent Huguenots 1700 | 1.4936*** | 1.4936*** | 1.4936 |
| | (0.5097) | (0.2324) | (1.5165) |
| Intercept | Yes | Yes | Yes |
| Additional Controls | Yes | Yes | Yes |
| Observations | 597 | 597 | 597 |
| Number of Clusters/Towns | 250 | 250 | 250 |

| *Panel B* | (4) | (5) | (6) |
|---|---|---|---|
| | Cluster Bootstrap | Cluster Bootstrap | Cluster Bootstrap |
| | 50 Reps - Different Seed | 400 Replications | 2000 Replications |
| Percent Huguenots 1700 | 1.4936 | 1.4936 | 1.4936 |
| | (1.0930) | (1.6894) | (1.4354) |
| Intercept | Yes | Yes | Yes |
| Additional Controls | Yes | Yes | Yes |
| Observations | 597 | 597 | 597 |
| Number of Clusters/Towns | 250 | 250 | 250 |

*Notes:* Textile industry data is taken from Hornung (2014b). Regression specification is the same as in Hornung (2014a, 104) Table 2 Column (5). Cluster-robust standard errors are clustered on town level. Additional controls that are not displayed are: Not using looms (dummy), ln town population 1802, Merino sheep, p.c. 1816 (county), Percent Protestant 1816, Not Prussia in 1700 (dummy), Dummy for imputed values, Relevant textile production 1680. Standard errors are reported in parantheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## Table 4: Weak Instrument Statistics

| *Panel A* | (1) | (2) |
|---|---|---|
| IV Specification | Table 2 - (2, 3) | Table 2 - (4, 5) |
| Instrument Construction: | Unadjusted | Interpolated |
| Instrumented Variable: | PC Huguenots in 1700 | PC Huguenots in 1700 |
| First Stage $R^2$ | 0.5931 | 0.6005 |
| (Shea's) Partial $R^2$ | 0.1501 | 0.1656 |
| Cragg-Donald F statistic | 24.37 | 27.38 |
| Kleinbergen-Paap rK Wald F-Statistic | 3.67 | 4.79 |

| *Panel B* | (3) | (4) |
|---|---|---|
| IV Specification | Table 2 - (6,7) | Hornung (2014a, 110) T. 5 - (4, 5) |
| Instrument Construction: | Aggregated | Aggregated |
| Instrumented Variable: | PC Huguenots in 1700 | ln Huguenots in 1700 |
| First Stage $R^2$ | 0.6108 | 0.7537 |
| (Shea's) Partial $R^2$ | 0.2407 | 0.2994 |
| Cragg-Donald F statistic | 55.15 | 74.37 |
| Kleinbergen-Paap rK Wald F-Statistic | 5.74 | 15.35 |

*Notes:* Textile industry data is taken from Hornung (2014b). Table reports first-stage IV diagnostic statistics. Sargan statistic for overidentification test of all instruments is not reported because the model is exactly identified (one endogenous regressor and one instrumental variable candidate).

Table 5: OLS With Different Imputation Methods For Ln Input 1802

| Dependent Variable: ln output 1802 | (1) MI Hornung | (2) Mean | (3) Random | (4) Regression | (5) Iterative Reg. | (6) Random Reg. |
|---|---|---|---|---|---|---|
| Percent Huguenots 1700 | 1.4588 | -0.7063 | -0.5629 | 1.7316 | 0.2748 | 1.6577 |
| *Cluster-Robust* | (0.2100)*** | (0.9109) | (0.8281) | (0.6987)** | (0.8799) | (0.6926)** |
| *Cluster-Bootstrap* | (0.5397)*** | (3.2549) | (2.6776) | (1.8797) | (3.5104) | (1.9278) |
| Dummy for imputed values | 0.0323 | 0.6187 | 0.3902 | 0.2127 | 0.4047 | 0.2100 |
| *Cluster-Robust* | (0.0342) | (0.1175)*** | (0.1268)*** | (0.1051)** | (0.1280)*** | (0.1046)** |
| *Cluster-Bootstrap* | (0.0357) | (0.1196)*** | (0.1383)*** | (0.1457) | (0.1359)*** | (0.1350) |
| ln input 1802 (MI Hornung) | 0.8054 | | | | | |
| *Cluster-Robust* | (0.0219)*** | | | | | |
| *Cluster-Bootstrap* | (0.0214)*** | | | | | |
| ln input 1802 (mean imputed) | | 0.5505 | | | | |
| *Cluster-Robust* | | (0.0482)*** | | | | |
| *Cluster-Bootstrap* | | (0.0472)*** | | | | |
| ln input 1802 (random imputed) | | | 0.3929 | | | |
| *Cluster-Robust* | | | (0.0442)*** | | | |
| *Cluster-Bootstrap* | | | (0.0470)*** | | | |
| ln input 1802 (reg. imputed) | | | | 0.0640 | | |
| *Cluster-Robust* | | | | (0.0502) | | |
| *Cluster-Bootstrap* | | | | (0.3059) | | |
| ln input 1802 (iterative reg. imp.) | | | | | 0.3450 | |
| *Cluster-Robust* | | | | | (0.0535)*** | |
| *Cluster-Bootstrap* | | | | | (0.0547)*** | |
| ln input 1802 (random reg. imp.) | | | | | | 0.0635 |
| *Cluster-Robust* | | | | | | (0.0490) |
| *Cluster-Bootstrap* | | | | | | (0.2556) |
| Intercept | Yes | Yes | Yes | Yes | Yes | Yes |
| Additional Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 693 | 693 | 693 | 693 | 693 | 693 |
| $R^2$ | 0.9669 | 0.9077 | 0.8842 | 0.8276 | 0.8749 | 0.8275 |
| Adjusted $R^2$ | 0.9664 | 0.9062 | 0.8823 | 0.8248 | 0.8729 | 0.8247 |

*Notes:* Textile industry data is partly taken from Hornung (2014b). Around 15 percent of all input values are imputed. Table is comparable to Hornung's (2014, 104) Table 2 Column 6. Cluster-bootstrap standard errors are computed using 400 replications. Additional control variables not displayed are: Ln workers 1802, ln looms 1802, Not using looms (dummy), ln town population 1802, Merino sheep, p.c. 1816 (county), Percent Protestants 1806, Not Prussia in 1700 (dummy), Relevant textile production 1608. Standard errors are reported in parantheses: $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

# List of Appendices



# A Cluster-Bootstrap Algorithm

This section displays the general bootstrap algorithm following Cameron and Trivedi (2005, 360) and Efron and Tibshirani (1994, 47):

1. Given data $x_1,...,x_n$, draw B independent bootstrap samples $x^{1*},...,x^{B*}$, each consisting of n data values drawn with replacement from $x$.

   $\rightarrow$ Cluster-Bootstrap: Instead sample observations from different groups/clusters and proceed accordingly.

2. Calculate the statistic of interest such as the standard error $se(\hat{\theta}^*)$ of the estimator $\theta^*$ for one bootstrap sample.

3. Repeat step 2 for each of the bootstrap samples, obtaining B bootstrap replications of the statistic of interest: $se(\hat{\theta}^{1*}),...,se(\hat{\theta}^{B*})$.

4. Using these B bootstrap replications one can obtain the bootstrapped version of the standard error statistic in the following way:

$$se_{Boots}(\hat{\theta}) = \sqrt{\frac{\sum_{b=1}^{B} \hat{\theta}^*(b) - \bar{\hat{\theta}}^*(b)}{(B-1)}} \quad \text{where} \quad \bar{\hat{\theta}}^*(b) = \frac{\sum_{b=1}^{B} \hat{\theta}^*(b)}{B} \tag{A.1}$$

# B Consistency of IV Estimator

Following Dougherty (2011, 339), assume the following univariate model with a failure of the exogeneity assumption and a valid instument $z$:

$$y = \beta_1 + \beta_2 x + u \text{ with } Cov(x,u) \neq 0 \tag{B.1}$$

$$\hat{\beta}_2^{OLS} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})} \tag{B.2}$$

$$\begin{aligned}
\hat{\beta}_2^{IV} =& \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})} && \text{(B.3)}\\
=& \frac{\sum(z_i - \bar{z})[(\beta_1 + \beta_2 x_i + u_i) - (\beta_1 + \beta_2 \bar{x} + \bar{u})]}{\sum(z_i - \bar{z})(x_i - \bar{x})} && \text{(B.4)}\\
=& \frac{\sum(z_i - \bar{z})[\beta_2(x_i - \bar{x}) + u_i - \bar{u}]}{\sum(z_i - \bar{z})(x_i - \bar{x})} && \text{(B.5)}\\
=& \beta_2 + \frac{\sum(z_i - \bar{z})(u_i - \bar{u})}{\sum(z_i - \bar{z})(x_i - \bar{x})} && \text{(B.6)}
\end{aligned}$$

It is not possible to take expectations of this expression. Due to the fact that x is not independent of $u$ there is no closed-form expression of the fraction in (B.6). Therefore, one has to be content with the large-sample properties of the estimator (Dougherty, 2011, 339):

$$\begin{aligned}
plim(\hat{\beta}_2^{IV}) =& \beta_2 + plim\left[\frac{\sum_{i=1}^{n}(z_i - \bar{z})(u_i - \bar{u})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}\right] && \text{(B.7)}\\
=& \beta_2 + plim\left[\frac{\frac{1}{n}\sum(z_i - \bar{z})(u_i - \bar{u})}{\frac{1}{n}\sum(z_i - \bar{z})(x_i - \bar{x})}\right] && \text{(B.8)}\\
=& \beta_2 + \frac{Cov(z,u)}{Cov(z,x)} && \text{(B.9)}
\end{aligned}$$

By the exclusion restriction we have $Cov(z,u) = 0$ and by instrument relevance $Cov(z,x) \neq 0$. Therefore, we can show that the estimator is consistent.

$$plim(\hat{\beta}_2^{IV}) = \beta_2 + \frac{0}{\sigma_{z,x}} = \beta_2 \tag{B.10}$$

# References

**Acemoglu, Daron.** 2009. *Introduction to modern economic growth.*: Princeton University Press.

**Angrist, Joshua D., and Alan B. Krueger.** 2001. "Instrumental variables and the search for identification: From supply and demand to natural experiments." *Journal of Economic Perspectives*, 15(4): 69–85.

**Angrist, Joshua D., and Jörn-Steffen Pischke.** 2009. *Mostly harmless econometrics: An empiricist's companion.*: Princeton University press.

**Behre, Otto.** 1905. *Geschichte der Statistik in Brandenburg-Preußen bis zur Gründung des Königlichen Statistischen Bureaus.* Berlin: Carl Heymanns Verlag, Quoted in Hornung (2014a).

**Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan.** 2004. "How Much Should We Trust Differences-In-Differences Estimates?" *The Quarterly Journal of Economics*, 119(1): 249–275.

**Borjas, George J.** 2014. *Immigration economics.*: Harvard University Press.

**Borjas, George J., and Kirk B. Doran.** 2012. "The Collapse of the Soviet Union and the Productivity of American Mathematicians." *The Quarterly Journal of Economics*, p. 161.

**Bound, John, David A. Jaeger, and Regina M. Baker.** 1995. "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak." *Journal of the American statistical association*, 90(430): 443–450.

**Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller.** 2008. "Bootstrap-based improvements for inference with clustered errors." *The Review of Economics and Statistics*, 90(3): 414–427.

**Cameron, A. Colin, and Douglas L. Miller.** 2015. "A practitioners guide to cluster-robust inference." *Journal of Human Resources*, 50(2): 317–372.

**Cameron, A. Colin, and Pravin K. Trivedi.** 2005. *Microeconometrics: Methods and Applications.*: Cambridge University press.

**Cameron, A. Colin, and Pravin K. Trivedi.** 2009. *Microeconomics using stata.*: Lakeway Drive, TX: Stata Press Books, Revised Edition.

**Dougherty, Christopher.** 2011. *Introduction to econometrics.*: Oxford university press.

**Efron, Bradley, and Robert J. Tibshirani.** 1994. *An introduction to the bootstrap.*: CRC press.

**Gelman, Andrew, and Jennifer Hill.** 2006. *Data analysis using regression and multilevel/hierarchical models.*: Cambridge University Press.

**Hornung, Erik.** 2014a. "Immigration and the diffusion of technology: The Huguenot diaspora in Prussia." *The American Economic Review*, 104(1): 84–122.

**Hornung, Erik.** 2014b. "Immigration and the diffusion of technology: The Huguenot diaspora in Prussia: Data Set." *The American Economic Review*, 104(1): –, URL: `https://www.aeaweb.org/articles.php?doi=10.1257/aer.104.1.84`- Accessed February 15, 2016.

**Hornung, Erik.** 2014c. "Immigration and the diffusion of technology: The Huguenot diaspora in Prussia: Online Appendix." *The American Economic Review*, 104(1): –, URL: `https://www.aeaweb.org/aer/data/jan2014/20111335_app.pdf`.

**Kennedy, Peter.** 2013. *A Guide to Econometrics.*: Blackwell Publishing, 6th edition.

**Kerr, Sari Pekkala, and William R. Kerr.** 2001. "Economic Impacts of Immigration: A survey." *Finnish Economic Papers*, 24(1): 1–32.

**Keyser, Erich.** 1939-1974. *Deutsches Städtebuch-Handbuch städtischer Geschichte*. 1-5, Stuttgart: Kohlhammer, Quoted in Hornung (2014a).

**Krug, Leopold.** 1805. *Betrachtungen uber den Nationalreichtum des preussischen Staates und uber den Wohlstand seiner Bewohner*. 2: J. F. Unger, Quoted in Hornung (2014a).

**Lausberg, Michael.** 2007. *Hugenotten in Deutschland. Die Einwanderung von französischen Glaubensflüchtlingen*. Marburg: Tectum Verlag.

**Moser, Petra, Alessandra Voena, and Fabian Waldinger.** 2014. "German Jewish Emigres and US Invention." *American Economic Review*, 104(10): 3222–55.

**Moulton, Brent R.** 1986. "Random group effects and the precision of regression estimates." *Journal of econometrics*, 32(3): 385–397.

**Pfister, Christian.** 2007. *Bevölkerungsgeschichte und historische Demographie: 1500-1800*. 28: Oldenbourg Verlag, Quoted in Hornung (2014a).

**Portes, Alejandro.** 1995. *The economic sociology of immigration: Essays on networks, ethnicity, and entrepreneurship.*: Russell Sage Foundation.

**Rubin, Donald B., and Roderick Little.** 2002. *Statistical analysis with missing data.*: J Wiley & Sons.

**Schmoller, Gustav.** 1922. *Deutsches Stadtewesen in älterer zeit.*: Kurt Schroeder, Quoted in Hornung (2014a).

**Shea, John.** 1997. "Instrument relevance in multivariate linear models: A simple measure." *Review of Economics and Statistics*, 79(2): 348–352.

**Staiger, Douglas, and James H. Stock.** 1997. "Instrumental Variables with Weak Instruments." *Econometrica*, 65(3): 557–586.

**Verbeek, Marno.** 2008. *A guide to modern econometrics.*: John Wiley & Sons.

**Waldinger, Fabian.** 2010. "Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany." *Journal of Political Economy*, 118(4): 787–831.

**Williams, Allan, and Vladimír Baláž.** 2008. *International migration and knowledge.*: Routledge.

**Wohlfeil, Rainer.** 1976. *Kriegsverlauf 1635 bis 1642: Bevölkerungsverluste der brandenburgischen Städte zwischen 1625 und 1652/53, (Der Dreißigjährige Krieg II).* Berlin: de Guyter, Quoted in Hornung (2014a).

**Wooldridge, Jeffrey M.** 2010. *Econometric analysis of cross section and panel data.*: MIT press.

**Wooldridge, Jeffrey M.** 2012. *Introductory econometrics: A modern approach.*: Cengage Learning.