

Discussion  
Papers

Deutsches Institut für Wirtschaftsforschung

2018

An Offer that you Can't Refuse?  
Agrimafias and Migrant Labor on  
Vineyards in Southern Italy

Stefan Seifert and Marica Valente

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

#### IMPRESSUM

© DIW Berlin, 2018

DIW Berlin  
German Institute for Economic Research  
Mohrenstr. 58  
10117 Berlin

Tel. +49 (30) 897 89-0  
Fax +49 (30) 897 89-200  
<http://www.diw.de>

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:  
<http://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:  
<http://ideas.repec.org/s/diw/diwwpp.html>  
<http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

An offer that you can't refuse?  
Agrimafias and Migrant Labor on Vineyards  
in Southern Italy

Stefan Seifert\* and Marica Valente†

**Abstract** In the 2011 post-Arab Spring migration wave, over 64,000 migrants landed on the southern Italian coast, with many of them potentially working illegally on farms through *caporalato*, a widespread system of illegal recruitment of underpaid farm labor run by Italian agrimafias. To test this hypothesis, this paper evaluates the causal effects of the 2011 migration wave on reported labor productivity focusing on vineyards in southern Italy. Based on a dynamic panel data model, labor productivity is estimated to increase by about 11% on average for 2011 and 2012. We show that this corresponds to a total of around 10 million unreported work hours, or 21,000 full-time employees, in each year. We interpret this as an increase in employment of illegal workforce due to the migration wave. Magnitude, direction, and statistical significance of the effect are confirmed under various model specifications and using synthetic control and post-lasso approaches.

**Keywords:** Migration wave, Agrimafias, Illegal Employment, Vineyard productivity

**JEL Codes:** F22, J61, J43

---

\*Corresponding Author. University of Bonn, Institute for Food and Resource Economics, Bonn, Germany. Email: s.seifert@ilr.uni-bonn.de

†Humboldt University of Berlin and DIW Berlin – German Institute for Economic Research, Berlin, Germany. Email: mvalente@diw.de

# 1 Introduction

In 2011, in the aftermath of the Arab Spring uprisings and the escalating Syrian civil war, the largest migration wave of the last decades crossed the central Mediterranean, with over 64,000 migrants landing in the southern Italian regions of Apulia, Calabria and, primarily, Sicily (FRONTEX, 2016). After coming ashore, migrants usually want to continue onwards to other EU countries like Germany and Sweden. However, any undocumented migrant landing on Italian coasts is irregular and at risk of expulsion (Italian Penal Code, 2009). Relying on the promise to obtain (forged) documents and facilitate journeys, migrants enter *caporalato*, an increasingly widespread system run by agrimafias that recruit and exploit underpaid workforce (Flai-Cgil, 2016). However, although up to 500,000 immigrants are estimated to be irregularly employed as crop farm workers in Italy (Flai-Cgil, 2014), there is little empirical evidence on the impact of migration waves on agricultural labor markets, in particular in the European Union (EU).<sup>1</sup>

For the first time, this study investigates potential illegal employment of migrant labor in farmlands of southern Italy after the 2011 migration wave. This large and unexpected migrant inflow characterizes a natural experiment (Dustmann et al., 2016b; Peri, 2016) that we use to estimate causal effects on labor productivity in the grape growing business, assuming that illegal employment would lead to overreported labor productivity. We focus on Sicily and Apulia, which are the recipients of the migrant wave (FRONTEX, 2016) and among Italy's major players for wine and grape production.

Looking at illegal labor employment in the wine and grape sectors of Sicily and Apulia is particularly meaningful due to the increasing competition across the Italian and southern European markets, and especially: (I) The strong rise of international wine production between 2000 and 2010, and the arrival of non-EU wines; (II) shrinking demand globally for wine and a credit freeze following the 2008 economic crisis; (III) climatic shocks, especially in 2002; and (IV) failed agricultural policies leading to deregulation of the EU wine sector (CMO, 2008). These developments demanded new strategies to cut costs and increase international competitiveness, in particular for the (small) less favored grape growing areas at risk of further depression and possible abandonment (Gaeta and Corsinovi, 2014). In parallel, Italian agrimafias supply underpaid illegal workforce to farmers, e.g., for grape harvesting, which may allow farmers to reduce factor costs to sustain the increasing global

---

<sup>1</sup>The estimated value of agrimafias' illegal business in Italy ranges from 14 to 17.5 billion euros, with profits earned from trading crops and refined goods like tomatoes, grapes, wine, and animal-derived products. Of the 10,311 firms sequestered or confiscated in Italy between 1982 and 2016, about 50% are farms, and 75% were detected only since 2011 (Flai-Cgil, 2016).

competition. This, however, comes at the cost of illegal migrant workers who lack any legal protection, cannot leave the illegal framework, and are heavily exploited (Flai-Cgil, 2016). To identify such illegal employment, we assume that labor productivity measures increase because the illegal share of labor input is not reported. The major transmission channel affecting labor productivity is a “displacement effect,” in which illegal labor can be employed to replace legal labor inputs. This results in an increase in reported labor productivity under the assumption that illegal labor can substitute homogeneous unskilled labor flexibly, and without sacrificing learning effects. Further, due to the high labor intensity and low mechanization of the grape growing sector of the treated regions, a displacement effect of capital with illegal labor is rather unlikely.

Labor market effects in the grape growing sector are expected in immediate response to both agrimafias’ activity and landed migrants with high incentives to enter illegal labor channels to quickly obtain documents and/or journey facilitations. For this reason, labor market adjustments are likely to occur in the very short-term, when no reaction of other economic factors is expected. Furthermore, the 2011 migration wave, hereafter also called treatment, is an exogenous shock: It was unexpected and abnormally large, only hitting the southern Italian coasts due to their geographic vicinity. Thus, the treatment assignment can be considered to be random, which ensures independence between the treatment and potential outcomes. In addition, other political and economic dynamics of the immigrants’ countries of origin do not affect southern Italy as these economies are not interconnected and no spillovers occur. Lastly, no treatment effect is expected outside Sicily and Apulia because both agrimafias and a sudden migrant inflow are simultaneously present only in these regions.<sup>2</sup>

The empirical analysis uses farm-level data aggregated at the regional level for Italy and France between 1999 and 2012. Causal effects are analyzed using a Difference-in-Differences (DiD) framework in a dynamic linear panel regression model. Thereby, using a lagged dependent variable as an exogenous predictor allows accounting for the persistence characterizing labor productivity in the grape growing sector while simultaneously controlling for time-varying effects of unobserved labor productivity heterogeneity. Various tests support this model, and indicate that fixed effect approaches are likely not suited to account for the underlying dynamics in the data. As a robustness check against the residual presence of unobserved fixed (time-invariant) effects in the dynamic model specification, we implement an Anderson-Hsiao type regression (Anderson and Hsiao, 1981). To

---

<sup>2</sup>Calabria, region located between Sicily and Apulia, is excluded from the analysis due to missing data, but also it has only a very small wine sector.

further validate our results, we use the Synthetic Control Method (SCM) as a complementary (non-regression-based) approach following Abadie and Gardeazabal (2003), Cavallo et al. (2013), as well as Acemoglu et al. (2016) who extend the method for multiple treated units. Finally, we check for model misspecification and flexible functional form, and we estimate the causal effect on the selected model using the post-lasso estimator as introduced by Belloni et al. (2012, 2013).

Our results point toward an increase of illegal employment on vineyards after the 2011 migration wave. Indeed, we find that this labor supply shock has a statistically significant causal effect on labor productivity on southern Italian vineyards, and labor productivity increases on average by around 11% for the post-treatment period (2011 and 2012). Dynamic treatment effects indicate that the effect was stronger in 2012 (13%) than in 2011 (10%). We show that this effect corresponds to around 10 million hours irregularly worked in the treated regions in each year – or around 21,000 unreported full time employees. Given the absence of any technological, price, or additional labor market shocks in the analyzed period, our findings are likely explained by unreported employment of illegal workforce caused by the sudden 2011 inflow of migrants.

The remainder of the paper is organized as follows: Section 2 summarizes background information about this study and reviews the relevant literature. Section 3 outlines our estimation framework. The data are described in Section 4. Section 5 presents the results, and Section 6 concludes.

## **2 Background and Literature**

### **2.1 Agrimafias and Migrant Labor in Southern Italy**

Italy’s agricultural production is represented by medium and small-sized farms. To enable Italian agriculture to sustain competition with foreign markets, illegal migrant labor is highly exploited with wages below legal minimum thresholds (850 euros a month, about 5 euros per hour), averaging 40% lower than domestic wages (Joint ETI, 2015). The Italian Association for Juridical Studies on Immigration (ASGI, 2015) reports numerous violations of the EU Directive on illegal immigration and illegal employment of migrants without the required legal status in the EU (2009/52/EC). About 400,000 workers are estimated to be at risk of exploitation by caporalato, of which 80% are migrants (Flai-Cgil, 2014). Indeed, about a third of total agricultural employment, and up to 70% in Apulia at local level, is illegal (Eurispes, 2014).

The exploitation of migrant workforce is inevitably linked to illicit economic activities and, precisely, to the mafia which is, at its core, “a specific economic enterprise, an industry which produces, promotes and sells private protection” (Gambetta 1993, p.1). In the agricultural sector, the mafia mostly operates through the intermediation of gangmasters, called *caporali*, who negotiate with farmers and supply migrant workers. Caporali charge fees for workers’ transportation, food, phone charging, and accommodation, keeping about half of a worker’s daily salary. The latter consists of no more than 30 euros, with an hourly wage between 1.60 and 3 euros per hour over a 12 to 16-hour working day (Palmisano and Sagnet, 2016). However, for most migrants caporalato is often the only option to find “protection,” which includes a job, and, in the long-run, a residence permit.

Sicily and Apulia are characterized by the high density of both migrant landings and agrimafias controlling farmlands’ labor supply (Bandiera, 2003; FRONTEX, 2016). Once landed, migrants can enter agrimafias’ illegal labor channels through several mechanisms. After landing, migrants are identified and obtain first aid in temporary refugee camps called hotspots. However, migrants often want reach other EU countries like Germany and Sweden, which is difficult given their initial illegal status. Thus, many of them refuse identification, leave the hotspots and, with the aim to earn money and obtain documents, enter illegal labor and/or criminal channels (Dustmann et al., 2016a).<sup>3</sup> As a consequence, refugees and migrants are vulnerable to mafia organizations promising journey facilitations and (forged) documents offered in exchange for illegal labor in agriculture (EUROPOL-EMPACT, 2013; EUROPOL-EMSC, 2016).

Asylum seekers not refusing identification and waiting for the acceptance of their request, which may take more than two years, enter the System for the Protection of Asylum Seekers and Refugees (SPRAR) managed by local associations. The asylum seeker status, however, prevents legal access to the labor market during most of the waiting time. Further, due to high rejection rates of asylum requests (around 40% in 2011 and 60% in 2016; Italian Interior Ministry, 2016), migrants again have incentives to leave the facilities, look for an illegal job, and possibly obtain a work visa through their employer. In case of a failed asylum request, rejected asylum seekers and migrants are placed in detention centers, where they wait up to 18 month for the expulsion sentence, usually in very poor sanitary and living conditions (Human Rights Watch, 2014). Again, many immigrants leave detention centers and become potential illegal labor force (Corriere delle Migrazioni,

---

<sup>3</sup>By law, migrants without documents and asylum seekers are sent back to the EU member state where they first entered, as the latter is primarily responsible for assessing the asylum claim and/or the expulsion decision (Dublin Regulation, 2013). As a result, in 2015 one out of three landed migrants refused identification (Italian Court of Inquiry on Migration, 2015).

2015, MigrantSicily, 2016). In summary, agrimafias and migrant labor in agriculture are interlaced, moved by different, but reciprocal, interests.

## 2.2 Literature

There is a vast literature on the impacts of immigration on productivity and labor market outcomes that has developed (for overviews see Okkerse, 2008; Peri, 2016). These analyses are performed, e.g., for the US (e.g., Peri, 2012), the EU (e.g., Moreno-Galbis and Tritah, 2016), and also Italy (e.g., Venturini and Villosio, 2008; De Arcangelis et al., 2015). Very heterogeneous findings show both positive and negative effects on native wages, employment, and productivity (see Dustmann et al., 2016b, and citations therein). However, most studies focus on the long-term consequences of immigration rather than the short-term impact of sudden migrant inflows. Such a sharp and unexpected migrant inflow characterizes a positive labor supply shock and, as in our case, can also be seen as a natural experiment. The literature uses such settings to evaluate the effects of migration waves in several countries. One of the most analyzed cases is the 1980 Mariel boatlift, a mass migration of Cubans to the US, especially to Miami, which occurred when Fidel Castro opened the borders for a short window of time. Card (1990) was the first to study the labor market responses to this sudden migration wave and to compare the labor market outcomes in Miami and control cities. Using subgroups of the population most likely to compete with the migrants on the labor market, results indicate that Mariel immigrants increased the Miami labor force by 7%, but even more in unskilled occupations. Based on the synthetic control method (SCM), the Mariel boatlift has been re-evaluated by Borjas (2015, 2016) and Peri and Yasenov (2018). While the former finds considerable negative wage effects for some subgroups of the workforce suggesting displacement, the latter do not, likely due to the use of different samples.

Several papers analyze the labor market impacts of the migration waves following the outbreak of the Syrian civil war in 2011. Focusing on Turkey, Ceritoglu et al. (2015) estimate negative causal effects on employment in a DiD framework. Likewise, Del Carpio and Wagner (2015), Tumen (2016), and Balkan and Tumen (2016) conclude that a strong displacement of natives by immigrants occurs, especially in the informal sector, and, at the same time, unskilled men benefit from increasing employment opportunities. However, Peri (2016) argues that the identification of these causal effects could be biased due to potential spillovers between the neighboring Syria and Turkey, preventing potential outcomes caused by forced migration to be disentangled from other labor market adjustments.



A last stream of studies is concerned with repatriation, including a study by Hunt (1992) on emigration from Algeria to France after Algerian independence in 1962, a study by Carrington and De Lima (1996) on the repatriation from Africa toward Portugal in the 1970s, and an analysis of Russian immigration to Israel after the collapse of the Soviet Union studied by Friedberg (2001). However, although these studies also analyze sudden migration inflows, there are several identification problems, as discussed by Okkerse (2008) and Peri (2016). As Peri (2016) argues, a major problem can be the non-exogenous distribution of immigrants in the host countries, i.e., migrants chose their destinations, leading to omitted variable bias.

### 3 Methodology

This section describes the identification strategy used to estimate the causal effect of the 2011 post-Arab Spring migration wave on illegal labor employment on vineyards in southern Italy. Causal effects are identified by analyzing labor productivity under the assumption that employment of illegal labor leads to underreported labor input, and, thus, overreported labor productivity.

In the agricultural sector, unskilled labor productivity is fairly constant over time once we control for farm inputs, and production environment, e.g., weather conditions (Lamouria et al., 1963; Jorgenson and Griliches, 1967). This persistence of outcomes may be due to time-invariant, unit-specific unobservables (fixed effects) and due to so-called state dependence. State dependence (Heckman, 1981), i.e., a dynamic outcome process in which past outcomes help predicting future outcomes, can be explained by observed past behavior, but also by time-varying effects of unobserved labor productivity heterogeneity. The latter, if ignored, would cause inconsistent model estimates due to Omitted Variable Bias (OVB).

In particular, in our setting, the estimation of labor productivity would suffer from OVB if regional differences in farmers' propensity to hire illegal labor are not accounted for. This varies among regions and over time, and its determinants are, e.g., organized crime intensity, and farmers' attitudes toward risk and law obedience due to different incentive and monitoring systems at regional level. Further, the propensity to hire illegal labor in a region is likely serially correlated because farmers who have hired illegal labor in the past are more likely to do so also in the future.

To account for these effects, we model labor productivity as a dynamic linear process with limited memory under the assumption of sequential exogeneity. The latter means that

the inclusion of a Lagged Dependent Variable (LDV) makes the outcome conditionally independent on past values of observed and unobserved variables. In particular, the inclusion of a LDV controls for state dependence as long as it absorbs time-varying effects of unobserved labor productivity heterogeneity at regional level. On the contrary, a static model with fixed effects is not able to capture such state dependence. While combining both LDV and fixed effects is possible, testing for unobserved time-invariant, unit-specific effects allows to identify the most likely source(s) of this persistence (Breusch and Pagan, 1980; Honda, 1985).

The model to identify causal effects of the illegal labor supply shock writes:

$$y_{it} = \alpha k_{it} + \beta l_{it} + \gamma X_{it} + \delta D_{it} + \omega_{it} + \varepsilon_{it}, \quad (1)$$

where subscripts  $i$  and  $t$  indicate units and time, respectively,  $y_{it}$  is labor productivity,  $k_{it}, l_{it}$  are capital and land inputs, respectively,  $X_{it}$  are other exogenous regressors,  $\varepsilon_{it}$  are idiosyncratic unobserved shocks, and  $D_{it}$  is a treatment dummy. Following the potential outcome approach by Rubin (1974), the DiD estimator  $\hat{\delta}$  estimates the Average Treatment effect on the Treated (ATT) as the difference of two differences: the average outcome in the treatment group, i.e., Sicily and Apulia, before and after shock, and the average outcome in the control group before and after shock. Further,  $\omega_{it}$  is the time-varying effect of unobserved labor productivity heterogeneity that can be decomposed into dynamic and fixed components to absorb the persistence of the outcome:

$$\omega_{it} = \rho y_{it-1} + \mu_i + u_{it}, \quad (2)$$

where  $\mu_i$  are (unobserved) fixed effects as, e.g., soil quality, and  $y_{it-1}$  is the (observed) LDV capturing (unobserved) time-varying, unit-specific effects on  $y_{it}$  through the parameter  $\rho$ .<sup>4</sup>  $u_{it}$  are unobserved shocks to the current heterogeneity,  $\omega_{it}$ , that are not absorbed by the LDV.

If outcome persistence is mainly due to state dependence, the model includes only the dynamic component, i.e.,  $\omega_{it} = \rho y_{it-1} + u_{it}$ , and equation 1 becomes:

$$y_{it} = \rho y_{it-1} + \alpha k_{it} + \beta l_{it} + \gamma X_{it} + \delta D_{it} + \varepsilon_{it}^*, \quad (3)$$

with  $\varepsilon_{it}^*$  being the compound error term  $\varepsilon_{it} + u_{it}$  resulting from the substitution of (2) in (1).

---

<sup>4</sup>Any LDV model assumes two (tested) stability conditions on the autoregressive process, i.e., covariance stationarity ( $|\rho| < 1$ ) and weak dependence of the outcome (Hsiao, 2014).

The dynamic model 3 can be estimated by pooled OLS (DPOLS) under the assumption of exogenous regressors and sequential exogeneity. If the LDV absorbs the autocorrelation in the residuals, parameter estimates are unbiased and consistent.<sup>5</sup>

The identification of  $\delta$  in (3) is based on three assumptions: parallel trend (PTA), conditional independence (CIA), and stable unit treatment value assumption (SUTVA).

The PTA states that, conditional on the set of explanatory variables, the average outcomes of treated and untreated units follow the same trend. This assumption is, per se, untestable, as the shock deviates the trends. Following Autor (2003), we assess possible PTA violations in the pre-treatment period testing for joint significance of multiple in-time placebo treatment dummies.

The CIA requires that, conditional on explanatory variables, the assignment of the treatment is as good as random. Because the treated regions did not self-select into treatment and because regressors are chosen to be exogenous, this assumption is fulfilled in our case. Finally, SUTVA states that no spillover effects on the control units take place after the shock. As a safeguard against its violation, we restrict the analysis on the years until 2012, but exclude 2013. The reason is that landed migrants hosted in reception centers receive their expulsion or asylum decision within two years, after which they may escape or leave these facilities. At this point, they might also enter illegal labor channels in other regions (Giangrande, 2017), which may cause a violation of the SUTVA from 2013 onwards.

As a robustness check against the residual presence of unobserved fixed effects,  $\mu_i$ , we implement an Anderson-Hsiao type regression by first differencing the model and instrumenting the endogeneous first-differenced LDV with past outcome lags (AH, Anderson and Hsiao, 1981). First differencing may, however, induce new problems: First, if the DPOLS errors in (3) are not serially correlated, this induces autocorrelation in the first-differenced errors, weakening the exogeneity of the chosen lagged outcomes as instruments; second, first differencing may induce outcome cross-sectional dependence that was not present in levels.

As a complementary approach to address the OVB potentially caused by time-varying effects of unobserved labor productivity heterogeneity, we estimate the treatment effect using the SCM proposed by Abadie and Gardeazabal (2003). This counterfactual-based approach aims to estimate a statistical twin, aka synthetic control, of the treated unit against which the treatment effect is measured. This twin is derived as a convex com-

---

<sup>5</sup>If, instead, autocorrelation would still be present, the model is likely non-stationary even if  $|\rho| < 1$ , causing inconsistent and biased estimates (Keele and Kelly, 2006), and, in particular, upward bias for  $\hat{\rho}$ , and downward bias for the other coefficients (Achen, 2000).

bination of observations from the control group of untreated units using optimal weights estimated to resemble the treated unit in terms of outcome as well as outcome predictors (exogeneous covariates). The ATT is then derived by taking the average difference between the predicted (synthetic) and observed outcome of the treated unit over the post-treatment periods. This method may allow to derive an unbiased counterfactual also in the presence of unobserved time-varying, unit-specific effects if the counterfactual estimate fits the treated outcome and a set of outcome predictors over a relatively long pre-treatment period (Abadie and Gardeazabal, 2003; Abadie et al., 2010; Gobillon and Magnac, 2016). However, in the presence of observed, identifiable unit-specific effects in the pre-treatment period, such as weather shocks, this method is less suited to counterfactual estimation due to the impossibility to match upon binary variables. Despite this drawback, the SCM can be seen as a non-regression-based robustness check of the results obtained by DPOLS. Since the SCM was initially developed for settings with one treated unit only, we adjust the method according to our setting using averages of treated units as well as averages of separately estimated treatment effects as proposed by Cavallo et al. (2013) and Acemoglu et al. (2016).

To check for model misspecification and flexible functional form, and then estimate the ATT, we use the post-lasso regression method as introduced by Belloni et al. (2012, 2013). Thereby, we add other farm-related variables to the model, and we create many technical variables from the chosen covariates, in particular, interaction terms, log specifications, as well as second- and third-order orthogonal polynomials. We obtain a high-dimensional model with many parameters relative to the sample size that can be estimated under approximate sparsity, and, as before, conditional sequential exogeneity (for details see, e.g., Chernozhukov et al., 2013; Chernozhukov et al., 2017). For model selection, we apply lasso to the high-dimensional model, and, for ATT estimation, we apply DPOLS on the selected model (post-lasso). In the first step, lasso regularizes the regression by the penalized L1-norm to avoid overfitting, and, in the second step, post-lasso undoes the regularization bias caused by the parameter shrinkage.<sup>6</sup> By doing so, we aim to retain those variables that can increase prediction accuracy and, in other terms, reduce the variance of the predicted values. However, post-lasso model specifications lack theoretical justification and the economic interpretation of parameter estimates is often not straightforward. For this reason, we use post-lasso as a complementary approach to check if the ATT estimates by DPOLS is robust to alternative model specifications.

---

<sup>6</sup>Post-lasso is shown to perform at least as well as lasso in terms of the rate of convergence, and has the advantage of a smaller regularization bias (Belloni and Chernozhukov, 2013).

## 4 Data

We use data from the Farm Accountancy Data Network (FADN, European Commission, 2017) and explanatory variables from the Eurostat labor force survey (LFS, EUROSTAT, 2017). The dataset is a balanced panel over the period 1999-2012. Each cross-section is the average farm of each region on the *NUTS 2* level. The sample is representative due to stratified sampling and weighting. The farms in our sample are vineyards, i.e., farms specialized in grape and wine production.<sup>7</sup> The sample consists of 25 regions of which 14 are located in Italy and 11 in France, adding up to in total 350 observations.<sup>8</sup> These regions are comparable for several reasons. France and Italy share a border and have similar climatic conditions with warm Mediterranean climate in the south and temperate oceanic climate in the north. Further, both countries are EU member states and grape growers are working under mostly identical regulation. Finally, both countries have a long vitivincultural tradition.

We consider the two southern Italian regions Sicily and Apulia as the treated units, for which the treatment, i.e., the migration wave, takes place in 2011. As Figure 1 highlights, a sudden increase in landings of migrants on the southern Italian coast took place in the spring preceding the grape harvest of that year. Further, as reported by the local governments (Press Regione Puglia, 2011), a high share of young migrants escaped refugee camps and detention centers (e.g., 93% in Apulia) and might have been available as labor force from 2011 onwards.

However, despite data availability until 2013, we restrict the analysis to 2011 and 2012. The reasons are twofold. First, the effect of the treatment could be confounded by the additional landings registered in Sicily in 2013 (see Figure 1 and FRONTEX, 2016). Second, potential spillover effects from 2013 onwards may violate the assumption on the untreated status of the control regions (SUTVA). In fact, migrants that landed in 2011 and did not immediately flee from the hotspots in Sicily and Apulia are distributed among facilities all over Italy. At the earliest in 2013, immigrants receive their expulsion or asylum decision, after which they may escape or leave these facilities. At this point, they might also enter illegal labor channels in other regions (Giangrande, 2017), causing a violation of SUTVA. To analyze labor productivity, we define our logged dependent variable *LabProd* as the

---

<sup>7</sup>Additional farm income from other agricultural activities plays a minor role amounting to only 3.5% on average, with very low variation across regions and time.

<sup>8</sup>This selection excludes non-grape growing regions. For six regions single missing data points until 2003 are imputed using Multivariate Imputation by Chained Equations (MICE, see Buuren and Groothuis-Oudshoorn, 2011). Few extreme observations are also handled with transformation.

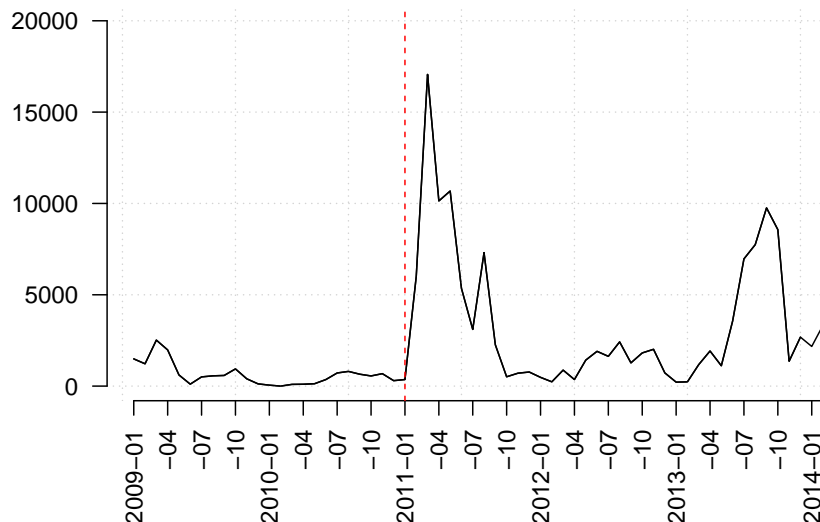


Figure 1: Boarder crossings to Italy via the central Mediterranean route  
(Own illustration, source: FRONTEX (2016))

total output from crops in euros divided by total hours worked, which is the sum of all paid and unpaid hours worked.

While this measure controls effectively for seasonal/unseasonal jobs and working regimes, it should be noted that it may vary due to quantity and price variations, with the latter being the main source of concern. However, we are confident in ruling out such effects for several reasons. First, price effects could be sizable only for high-quality wines, i.e., those with Protected Designation of Origin (PDO). In 2011, in Sicily (Apulian) PDOs account for only 4% (15%) of the regional wine production, and for only 1% (6%) of total Italian PDO production (Baccaglio, 2016a). Thus, price fluctuations can have only limited impact. Second, no sizable price shocks among Sicilian and Apulian PDOs have been registered. Third, although we observe price variations for single PDOs, average prices follow similar trends in all regions (Baccaglio, 2016b), with the only exception of Veneto that experienced the Prosecco-boom with a price and sales increase by around 50% between 2009 and 2012.

Our explanatory variables are chosen to be exogenous, and contain measures of capital, land, labor, as well as other farm and labor market characteristics. Regarding capital, we include capital intensity in terms of book values of machinery over total vineyard hectares (in logs,  $\ln Machinery$ ). In fact, the degree of capital intensity determines both workers'

day-long fatigue level and the efficiency of grower management practices, two main drivers of labor productivity in grape harvesting and grapevine pruning (Lamouria et al., 1963). For the same reason, we also account for the lagged investment rate computed as gross investment over total fixed assets (*invRate*). Land input is included as total vineyard hectares (in logs, *lnLand*) to capture potential returns to scale. Instead, to account for different degrees of vineyard specialization at regional level, we use the share of vineyard hectares over total utilized agricultural area (*vineSpec*). We also use two variables of labor market characteristics to control for potential competition on the labor market: the deviation of the regional unemployment rate from its long-term mean (*unempRate*), and the share of population above the age of 15 with less than primary or secondary education (ISCED11) (*unskilledLab*) as an indicator of unskilled workforce available for the harvesting. To account for weather effects that may impact both potential output and harvested output through workers' fatigue, we include the average of the daily minimum temperatures measured at the major regional airports (*minTemp*).<sup>9</sup> Three additional control variables are included: A fixed effect for France to account for unobserved systematic differences in labor productivity at country-level (*france*), a linear time trend (*trend*), and a time dummy to absorb the effects on both vineyard output and hours worked of 2002 anomalous weather events including heavy precipitations that destroyed considerable shares of the harvest in both Italy and France (*weather<sub>2002</sub>*). The treatment dummy,  $D_{it}$ , used to estimate the treatment effect,  $\delta$ , is the interaction term of *treated*  $\times$  *post*, where *treated* is a unit dummy equal to one for Sicily and Apulia, and *post* is a time dummy equal to one for the post-treatment periods.

The boxplots in Figure 2 shows the descriptive statistics of the variables averaged over the observation period. For comparability and visualization reasons, each variable is normalized by dividing it by its mean. Detailed descriptive statistics are reported in the Appendix, Table 6. The figure shows that all variable values for Sicily and Apulia are within the range of the whole dataset, and mostly around the sample mean. In particular, the dependent variable for the treated units lies between the first and the second quartile of the sample, which is especially important to satisfy the common support assumption of the SCM.

---

<sup>9</sup>This data is scraped from [www.wunderground.com](http://www.wunderground.com).

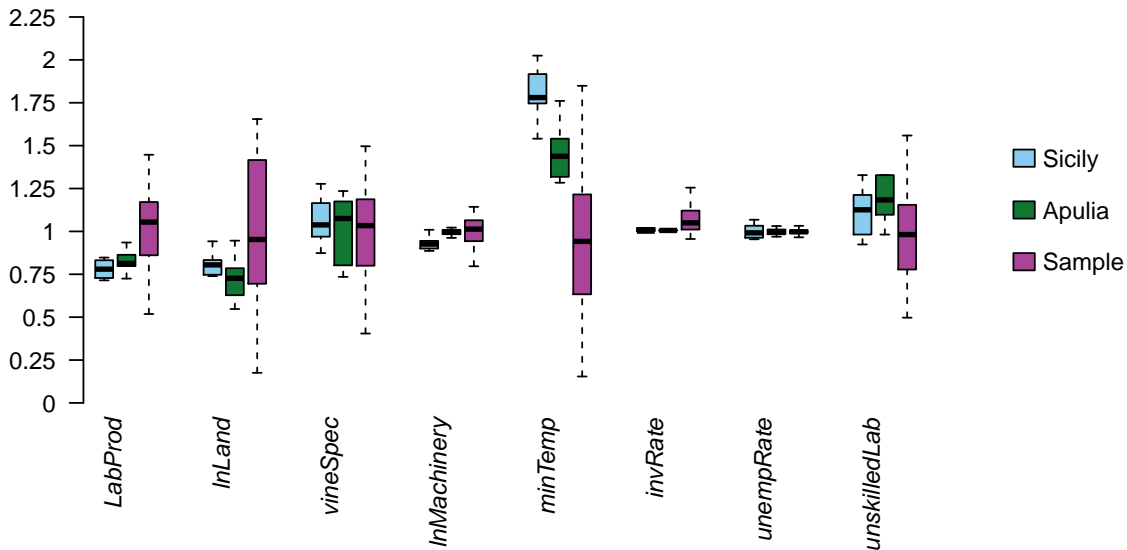


Figure 2: Descriptive statistics: Boxplots of mean-corrected variables

## 5 Results

We estimate the causal effects of the 2011 migration wave on vineyard labor productivity using a dynamic regression model as outlined in Section 3. After assessing the validity of our dynamic model specification, we estimate the DPOLS model in equation 3 with and without common time effects. Results from these models show that the 2011 migration wave led to a statistically significant increase in labor productivity of, on average, 11% in 2011 and 2012. To shed light on the evolution of illegal labor employment over time, dynamic treatment effects are also estimated and point toward negative adjustments of legal labor employment in the second year after the shock. Further, robustness checks by the Anderson-Hsiao type regression, the synthetic control and the post-lasso approach confirm the results outlined above.<sup>10</sup>

Table 1 reports the results of the DPOLS model. Column (1), (2), and (3) show these results for the DPOLS without time effects, with time effects (unit-demeaned), and with dynamic treatment effects, respectively. The DPOLS models with/without time effects show that the average causal effect estimate over 2011-2012, i.e., the ATT calculated as

<sup>10</sup>For the analysis, we use the statistical software R, and, particularly, the packages plm (Croissant and Millo, 2008), Synth (Abadie et al., 2011), and hdm (Chernozhukov et al., 2016).



	(M1)	(M2)	(M3)
$\rho$	0.714*** (0.040)	0.723*** (0.041)	0.723*** (0.041)
$\ln Land$	0.149*** (0.030)	0.136*** (0.029)	0.137*** (0.029)
$\text{vineSpec}$	0.018 (0.064)	-0.011 (0.059)	-0.011 (0.059)
$\ln Machinery$	0.119*** (0.028)	0.115*** (0.027)	0.115*** (0.027)
$\text{minTemp}$	-0.008 (0.005)	-0.006 (0.005)	-0.006 (0.005)
$\text{invRate}$	-0.191 (0.227)	-0.226 (0.237)	-0.228 (0.238)
$\text{unempRate}$	-0.924* (0.546)	-0.278 (0.663)	-0.284 (0.666)
$\text{unskilledLab}$	1.508** (0.604)	1.242** (0.604)	1.238** (0.605)
$\text{france}$	0.278*** (0.072)	0.272*** (0.072)	0.272*** (0.072)
$\text{weather}_{2002}$	0.133*** (0.036)		
$\text{trend}$	0.009** (0.004)		
$\text{treated}$	0.017 (0.037)	0.008 (0.036)	0.009 (0.036)
$\text{post}$	0.074** (0.037)		
$\delta$	0.102** (0.040)	0.111*** (0.038)	
$\delta_{2011}$			0.098** (0.048)
$\delta_{2012}$			0.123*** (0.043)
constant	-0.916*** (0.342)		
Time effects	No	Yes	Yes
F Statistic	450.2***	554.5***	515.3***
Adjusted R <sup>2</sup>	0.918	0.914	0.914
Notes:	***p=.01; **p=.05; *p=.1		

Table 1: DPOLS model without time effects (M1), with time effects (M2), and with dynamic treatment effect (M3)

$\exp(\hat{\delta}) - 1$ , ranges between 10.7% and 11.7%, and it is statistically significant at 1%. Considering the average causal estimate from these two models as its lower- and upper-bound, respectively, this means that the 2011 migration wave led to an abnormal average increase in *LabProd* of about 11.2%. Decomposing the ATT into yearly causal effects for 2011,  $D_{2011}$ , and 2012,  $D_{2012}$ , results indicate that the stronger effect occurs in 2012 with an increase of 13.1%. This seems plausible as 2011 saw other minor migration waves throughout the summer that further increased the illegal labor availability. Consequently, incentives to hire illegal labor likely increased after 2011, with negative adjustments of the legal labor demand.

The other explanatory variables show the expected signs. *lnLand* and *lnMachinery* indicate that more land and capital assets increase *LabProd*, pointing toward economies of scale. On the contrary, larger deviations from the long-term unemployment rate, *unemp*, are negatively related to *LabProd*. Given the systematic higher unemployment rate in the south of both France and Italy, as well as the relatively low variance of such variable over time, this partial effect may simply reflect the north-south gap in labor productivities. A higher availability of unskilled labor positively correlates with *Labprod*, likely indicating the higher labor productivity of more developed regional low-skilled labor markets. Lastly, *france* indicates that *Labprod* is on average higher in France than in Italy. The parameter for the 2002 extreme weather events also has the expected sign and positively affects *Labprod*, mainly due to the drop in labor force needed on vineyards in that year. Lastly, the autoregressive parameter  $\rho$  is statistically significant and amounts to about 0.7, confirming the presence of a well-behaved autoregressive process with a relatively high degree of state dependence.<sup>11</sup>

To establish the validity of the models and results outlined above, we conduct a series of tests. First, we test a static fixed effect (within) model for the residual presence of unobserved unit-specific effects by means of LM tests (Breusch and Pagan, 1980; Honda, 1985). Results indicate a significant, large residual variance across units (all p-values < 0.01), i.e., the presence of leftover unobserved unit-specific effects also after time-demeaning the model. Moreover, tests show that serial correlation of residuals is present (Breusch, 1978; Godfrey, 1978). Second, we perform the same tests on the dynamic model (equation 3). In this case, the lagged dependent variable absorbs these unobserved unit-specific effects (p-values > 0.1) and residuals' first-order serial correlation. Thus, parameter estimates

---

<sup>11</sup>The key stability conditions of the autoregressive outcome process are initially tested, and are fulfilled, i.e., the outcome covariance structure and unit-root tests indicate weak dependence and covariance stationarity, respectively (Choi, 2001; Hadri, 2000).

of the DPOLS model are consistent. Moreover, no cross-sectional dependence is detected (p-values  $> 0.06$ ) (Pesaran, 2004). Additionally, Newey-West corrected standard errors (Newey and West, 1994) account for (decaying) higher-order serial correlation.

Next, we transform our estimates on labor productivity into estimates of the unreported (illegal) hours worked at Sicily's and Apulia's vineyards. To do so, we perform simple back of the envelope calculations in the following way. Total production *Output* is a function of labor productivity *LabProd* and labor input *L* such that  $Output = LabProd^l * L^l + LabProd^{il} * L^{il}$  with superscripts *l* and *il* denoting legal and illegal labor input, respectively. Solving for  $L^{il}$  delivers the illegal input as a function of the observed values of output and legal labor input, while the values of true labor productivity of legal and illegal input are not observed. An estimate of the true labor productivity of legal input is calculated from the reported labor productivity  $LP^{reported}$  using the estimated average treatment effect:  $Lab\hat{P}rod^l = e^{-\hat{\delta}} * LabProd^{reported}$ . Further, we parameterize labor productivity of illegal input as a function of legal inputs as  $LabProd^{il} = \theta LabProd^l$ . This delivers:

$$\hat{L}^{il} = \frac{L^l \times (LabProd^{reported} - Lab\hat{P}rod^l)}{\theta Lab\hat{P}rod^l} = \frac{L^l (LabProd^{reported} [1 - e^{-\hat{\delta}}])}{\theta e^{-\hat{\delta}} * LabProd^{reported}} \quad (4)$$

The rationale behind this parameterization is that the productivity of illegal labor might differ from that of legal labor, e.g., due to fatigue from long working hours (Palmisano and Sagnet, 2016; Lamouria et al., 1963). Therefore,  $\theta$  relates the two labor productivities such that illegal labor and legal labor are identically productive for  $\theta = 1$ , and for example  $\theta = 1.3$  (0.7) indicates 30% higher (lower) labor productivity for illegal labor. Further, not only is the relationship between  $\hat{L}^{il}$  and  $\theta$  non-linear, but  $\hat{L}^{il}$  decreases faster for low productivity levels than for higher levels ( $\frac{\partial^2 \hat{L}^{il}}{\partial \theta^2} > 0$ ).

Table 2 reports the estimates of unreported work hours for the average farm, the three different model specifications, and for different values of  $\theta$  arbitrarily chosen to vary between 0.7 and 1.3. Estimates are calculated under the assumption of homogeneous treatment effects, i.e.,  $\delta$  is constant across the treated units. Estimates vary between years and regions. We find generally slightly higher values for Apulia than for Sicily, although differences are small. For identical productivity of legal and illegal workforce ( $\theta = 1$ ), unreported labor input is estimated to vary between 205 and 278 hours. These estimates correspond to 11% (205h/1800h) and 15% (278/1800) of an annual work unit, which is the agricultural equivalent of a full-time employee, as defined by the EU (European Commission, 2017), – or one person working around 5 to 7 40-hour weeks.<sup>12</sup> This estimate corresponds to the

---

<sup>12</sup>However, it should be noted that for illegal labor input, and in particular for labor input employed

length of the harvesting season, thus about one full-time worker might not be reported during this time.

While these estimates seem to be rather low, it should be noted that the sector is characterized by a strong fragmentation with a large number of fairly small vineyards. To estimate the overall effect, we transform our estimates at a regional level by multiplication with the number of farms within a region. Table 3 summarizes the results. For  $\theta = 1$ , total unreported labor input ranges between 9.0 and 10.3 million hours in 2011, and 10 and 12.2 million hours in 2012. Annual estimates range between 13 and 17.4 million hours for  $\theta = 0.7$ , and between 7.2 and 9.3 million hours for  $\theta = 1.3$ . Assuming 8-hour work days and a 60-day harvesting season, 10 million unreported hours correspond to around 21,000 illegally employed workers. While this corresponds to about one third of the 2011 migrant inflow, it is not clear to which degree this labor force consists of migrants and/or natives.

However, this calculation is sensitive to various factors: First, hours worked per day might be considerably higher. Second, estimates vary strongly with  $\theta$ . Thirdly, the estimated average treatment effect,  $\hat{\delta}$ , identifies only the increase in labor productivity due the migration wave. If labor productivity is already overestimated before this shock due to unreported labor, the estimated unreported hours are only a lower bound of the actual numbers, which is likely the case due to the historical presence of agrimafias (Flai-Cgil, 2016).

## 5.1 Robustness checks

### 5.1.1 In-time placebo DiD

To test for violation of DiD model assumptions, we introduce placebo treatment dummies for the five years before the shock. The left graph of Figure 3 displays the evolution of the placebo treatment effects over time (x-axis), and the respective parameter estimates as well as confidence intervals (y-axis). For 2011 and 2012, the estimated average causal effects result to be the largest in magnitude and with the smallest confidence levels. Joint statistical insignificance of placebo effects is not rejected with a p-value = 0.99. On the right, instead, parameter estimates (x-axis) are plotted against their respective standard errors (y-axis), and the shaded area indicates that there are no causal effects more extreme than those for 2011 and 2012.

---

through agrimafias, working conditions are usually tough, and 40-hour weeks are likely an underestimate.

	$\theta = 0.7$	$\theta = 0.8$	$\theta = 0.9$	$\theta = 1$	$\theta = 1.1$	$\theta = 1.2$	$\theta = 1.3$
<i>M1</i> ( $\hat{\delta} = 0.102$ )							
Sicily 2011	306	268	238	214	195	178	165
Apulia 2011	326	286	254	229	208	190	176
Sicily 2012	284	249	221	199	181	166	153
Apulia 2012	326	285	254	228	208	190	176
<i>M2</i> ( $\hat{\delta} = 0.111$ )							
Sicily 2011	334	293	260	234	213	195	180
Apulia 2011	357	312	278	250	227	208	192
Sicily 2012	311	272	242	217	198	181	167
Apulia 2012	357	312	277	250	227	208	192
<i>M3</i> ( $\hat{\delta}_{2011} = 0.098, \hat{\delta}_{2012} = 0.123$ )							
Sicily 2011	293	257	228	205	187	171	158
Apulia 2011	313	274	243	219	199	183	169
Sicily 2012	346	303	269	242	220	202	186
Apulia 2012	398	348	309	278	253	232	214

Table 2: Estimated unreported hours worked for the average farm by region

	$\theta = 0.7$	$\theta = 0.8$	$\theta = 0.9$	$\theta = 1$	$\theta = 1.1$	$\theta = 1.2$	$\theta = 1.3$
<i>M1</i> ( $\hat{\delta} = 0.102$ )							
Sicily 2011	6,658	5,825	5,178	4,660	4,237	3,884	3,585
Apulia 2011	6,930	6,064	5,390	4,851	4,410	4,043	3,732
Sicily 2012	7,417	6,490	5,769	5,192	4,720	4,327	3,994
Apulia 2012	6,926	6,061	5,387	4,848	4,408	4,040	3,730
<i>M2</i> ( $\hat{\delta} = 0.111$ )							
Sicily 2011	7,278	6,368	5,661	5,095	4,632	4,246	3,919
Apulia 2011	7,576	6,629	5,893	5,304	4,821	4,420	4,080
Sicily 2012	8,109	7,095	6,307	5,676	5,160	4,730	4,366
Apulia 2012	7,572	6,626	5,889	5,300	4,819	4,417	4,077
<i>M3</i> ( $\hat{\delta}_{2011} = 0.098, \hat{\delta}_{2012} = 0.123$ )							
Sicily 2011	6,384	5,586	4,965	4,468	4,062	3,724	3,437
Apulia 2011	6,645	5,814	5,168	4,652	4,229	3,876	3,578
Sicily 2012	9,041	7,911	7,032	6,328	5,753	5,274	4,868
Apulia 2012	8,442	7,387	6,566	5,909	5,372	4,925	4,546

Table 3: Estimated unreported hours worked by region (in thousands)

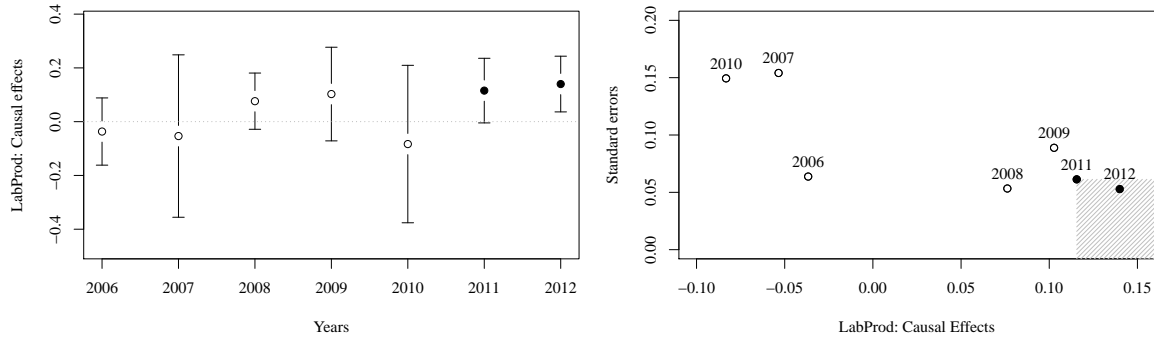


Figure 3: In-time placebos and dynamic causal effects

### 5.1.2 Anderson-Hsiao type regression

To further check for the robustness of our results against unobserved fixed effects,  $\mu_i$ , we use the AH estimator on the first-differenced dynamic model (Anderson and Hsiao, 1981). First differencing swaps away unit-specific fixed effects; however, it induces correlation between the first-differenced LDV and error term, thus, endogeneity. We choose the second and third lagged outcomes to instrument the first-differenced lagged outcome, both in levels and in first differences. Table 4 reports the results of the two-stage least square estimation using first-differenced instruments (column 1 and 2) and level instruments (column 3 and 4), respectively, with and without time effects. All four models show a positive causal effect of the migration wave on measured labor productivity. Parameter estimates of these causal effects are generally of higher magnitude than those obtained with the DPOLS specification (cp. Table 1, ranging between 15.4 and 17.9%). Despite the overall lower efficiency of the AH estimator, these estimates are statistically significant at the 1% level. The instruments are relevant according to robust F-tests on the first-stage regressions (p-values  $< 0.01$ ). Further, auxiliary-regression based Sargan tests with heteroskedasticity and autocorrelation corrected residuals indicate that the instruments are exogenous, i.e., uncorrelated with the AH model's residuals (p-value  $> 0.3$ ). However, it should be noted that first differencing the model induces serially correlated residuals, as well as unobserved unit-specific heterogeneity with a large residual variance across regions (p-values  $< 0.005$ ).<sup>13</sup> For this reason, both the exogeneity of the first-differenced lagged instruments and the model specification may be problematic. Therefore, we consider robustness of AH model estimates only as an indicator of robustness against alternative

<sup>13</sup>In this case, we follow Driscoll and Kraay (1994) and use standard errors corrected for both potential autocorrelation and cross-sectional dependence in the residuals.

model specifications.

	First-difference IVs		Level IVs	
	(AH1)	(AH2)	(AH3)	(AH4)
$\rho$	0.217 (0.189)	0.034 (0.132)	0.317 (0.270)	0.146 (0.154)
$\ln Land$	0.535*** (0.039)	0.517*** (0.048)	0.533*** (0.038)	0.517*** (0.046)
$\text{vineSpec}$	0.116 (0.241)	0.003 (0.275)	0.138 (0.246)	0.017 (0.300)
$\ln Machinery$	0.065* (0.039)	0.064* (0.036)	0.057 (0.043)	0.059 (0.037)
$\text{minTemp}$	0.004 (0.010)	0.013** (0.006)	0.003 (0.011)	0.013** (0.007)
$\text{invRate}$	0.249 (0.223)	0.200 (0.201)	0.217 (0.242)	0.171 (0.218)
$\text{unempRate}$	-0.336 (1.297)	-0.300 (2.041)	-0.319 (1.295)	-0.257 (2.117)
$\text{unskilledLab}$	-0.008 (0.845)	0.116 (0.995)	0.048 (0.915)	0.120 (1.075)
$\text{weather}_{2002}$	0.134*** (0.017)		0.127*** (0.023)	
$\text{post}$	0.026* (0.014)		0.025* (0.015)	
$\delta$	0.151*** (0.039)	0.143*** (0.029)	0.165*** (0.045)	0.161*** (0.032)
Time effects	No	Yes	No	Yes
F Statistic	9.199***	5.394***	8.346***	4.882***
Adjusted R2	0.151	0.190	0.122	0.099
<i>Notes:</i>	***p=.01; **p=.05; *p=.1			

Table 4: AH model estimates with different IV strategies and with/without time effects

### 5.1.3 Synthetic control approach

We implement the SCM as a complementary approach that captures time-varying effects of unobserved labor productivity heterogeneity. As the SCM was initially proposed to estimate treatment effects for only one treated unit, we follow two different strategies to apply the method in our setting. In a first approach, we create one artificial treated unit derived as the simple average of Sicily and Apulia. The counterfactual, i.e., a synthetic control unit against which the treatment effect is estimated, is derived by using an identical model specification as in the regression model M1. Using this approach, the treated unit's outcome and its counterfactual are very similar in the pre-treatment periods with

a correlation of over 80% and an  $R^2$  of around 60%. This suggests that the synthetic control captures time-varying effects of the treated-specific, serially correlated unobserved heterogeneity in labor productivity, which points toward unbiasedness of the synthetic control estimator. Post-treatment, the ATT, computed as the average post-treatment gap, amounts to 16.7%; thus, it has the same direction and similar magnitude to the ATT estimated with the DPOLS model.

To perform inference on the estimated ATT, we run in-space placebo tests and compute for each control unit a synthetic counterpart. Based on these placebo tests, empirical p-values for the estimated treatment effect can be calculated as the probability to obtain an ATT greater than or equal to the ATT of the treated unit (see Abadie et al., 2010, for details). Indeed, a higher treatment effect than for the average of Sicily and Apulia can be found for only one unit resulting in an empirical p-value of 7.7%.<sup>14</sup> Further, considering also the quality of the pre-treatment fit in terms of MSE, none of the placebos show a treatment effect of similar magnitude as for the average of Sicily and Apulia is found. Thus, placebo tests indicate that the 2011 migration wave had a statistically significant impact on labor productivity in the treated regions.

In a second approach, we estimate separate synthetic units for Sicily and Apulia to derive treatment effects, which are then averaged to obtain an estimate of the ATT. Estimation of both synthetic units is based on the same model as before. The resulting synthetic units indicate an average treatment effect of around 7% for Sicily and 19% for Apulia. However, the fit of synthetic and observed outcome paths are, in the pre-treatment periods, considerably better for Apulia. In particular, the synthetic unit for Sicily overestimates the observed outcome paths by around 11% which may result in an underestimation of the treatment effects.

To average these results, two approaches are available. First, Cavallo et al. (2013) suggest using simple averages of the treatment effects resulting in an overall ATT of 12.9%, thus of similar magnitude as the results obtained with DPOLS, but lower than the estimate obtained with the SCM for an averaged treated unit. Alternatively, following Acemoglu et al. (2016), one may use a weighted average in which the weights are obtained from the fit of synthetic and observed paths in the pre-treatment periods. Thereby, observations with a better fit before the treatment obtain a higher weight as these estimated treatment effects shall contain more information. Calculating the weights as  $W_i = 1 - MSE_i / (MSE_i +$

---

<sup>14</sup>We exclude the treated unit from placebo counterfactual estimation as it otherwise may induce bias in the estimated ATT. Additionally, counterfactuals with a pre-treatment MSE three times higher than the treated unit's MSE are discarded to safeguard against out-of-support estimations (Abadie et al., 2010).



$MSE_j$ ) with  $i$  and  $j$  denoting Sicily and Apulia delivers a weight of 75% (25%) for the effect estimated for Apulia (Sicily). Thus, using this approach, we obtain an average treatment effect of 15.8% that is higher than the effects obtained with DPOLS, but similar to those obtained with AH-type regression.

#### 5.1.4 Post-lasso approach

To check for possible model misspecification and incorrect functional form, we enlarge our main model (M1) with additional farm-related variables, as well as high-dimensional covariates, in particular, interaction terms, log specifications, and second and third order orthogonal polynomials. We perform, first, model selection and, then, post-selection estimation of the ATT by post-lasso approach as proposed by Belloni et al. (2012, 2013). Main results from the post-lasso regression are the following: (I) The model specification is dynamic ( $\rho$  is not shrunk to zero); (II) rigorous lasso selects 17 variables, including some of our original covariates, some of their interaction terms, and some newly included farm-related variables; (III) after model selection, post-lasso DPOLS estimation reports plausible signs, and additionally selects the rent paid for farm land and buildings and rental charges (in logs); and (IV) the treatment dummy is highly statistically significant (p-value = 0.015), amounting to 9.3% (for details, see Table 7 in the Appendix).

Table 5 presents a comparison of the ATT estimates obtained by the previous methods, and the post-lasso approach. Although smaller in magnitude, the ATT computed by post-lasso confirms the presence of a statistically significant average causal effect of the 2011 migration wave on vineyard labor productivity in Sicily and Apulia, indicating the presence of illegal labor employment.

Model	Estimate	Std. Error
DPOLS: M1	0.102	0.040
DPOLS: M2	0.111	0.038
AH2 (upper bound)	0.143	0.029
AH3 (lower bound)	0.165	0.045
SCM (Treated Unit)	0.167	-
SCM (Acemoglu et al.)	0.158	-
SCM (Cavallo et al.)	0.129	-
Post-lasso	0.093	0.037

Table 5: Comparison of ATT estimates

## 6 Conclusions

In this paper, we analyze whether the 2011 migration wave caused an increase in illegal employment. In particular, focusing on vineyards in the southern Italian regions of Sicily and Apulia, we estimate the causal effect of this large and unexpected labor supply shock on labor productivity using a representative sample of farm-level data aggregated at regional level for vineyards in Italy and France between 1999 and 2012. Identification of the causal effect is achieved through a difference-in-differences estimator in a dynamic panel regression framework that allows controlling for time-varying effects of unobserved labor productivity heterogeneity across regions. Alternative model specifications using an Anderson-Hsiao type regression are also employed. Finally, results are validated by synthetic control and post-lasso approaches.

Our results show that the 2011 migrant wave caused a statistically significant increase of labor productivity of around 11% in 2011 and 2012. When considering dynamic treatment effects, this effect is found to be stronger in 2012, although with small differences (13 vs 10%). We show that this effect corresponds to around 10 million hours irregularly worked in the treated regions in each year – or around 21,000 full-time employees. Thereby, illegal workforce may have displaced legal workforce, leading to underreported labor input and overreported labor productivity. Further, our findings suggest that agrimafias immediately responded to the labor supply shock matching the needs of irregular migrants looking for documents with those of grape growers facing an increasing competitive pressure. These results are in line with the literature (see Dustmann et al., 2016b; Peri, 2016), which finds low-skilled jobs the most vulnerable to migrant labor supply shocks. Indeed, this is the case for vineyard labor: The seasonal nature and the low skill requirements of field picker jobs limit workers’ bargaining power and makes them substitutable.

Several questions remain open and should be addressed in future research. Generally, the impacts of the post-Arab Spring migration crisis on European labor markets needs further investigation. In particular, in addition to employment effects, the impact of the migration wave on wages of both legal and illegal labor should be analyzed. Further, the current analysis should be extended to the whole agribusiness as it is the sector that absorbs most of the illegal workforce. Finally, long-term effects of the newest migration waves on labor markets need to be evaluated taking into account the current EU immigration policy and the recent regulatory efforts against labor exploitation.

## Acknowledgements

We thank the American Association of Wine Economists for financial support, and Mariusz Stefan Migas (European Commission, DG AGRI), Mauro Santangelo (CREA), and Tiziana Sarnari (ISMEA) for providing data. The paper greatly benefitted from discussions at the 2017 conferences of the American Association of Wine Economists (AAWE, Padua), the Agricultural Economics Society (AES, Dublin), the European Workshop on Efficiency and Productivity Analysis (EWEPA, London), and the Econometrics of Panel Data and Network Analysis Summer School (CRC, Berlin), as well as at the seminars of the Loughborough University, University of Verona, and DIW Berlin. We thank Pio Baake, Boris Bravo-Ureta, Tomaso Duso, Bernd Fitzenberger, Cristina Salvioni, Jeffrey Wooldridge, and Angelo Zago for fruitful discussion and input.

## 7 Appendix

	Sicily & Apulia				Control units			
	Mean	Min	Max	SD	Mean	Min	Max	SD
LabProd	2.441	2.170	2.840	0.179	3.089	1.574	4.395	0.599
lnLand	1.534	1.092	1.887	0.195	2.035	0.350	3.301	0.751
vineSpec	0.671	0.472	0.821	0.099	0.640	0.260	0.961	0.153
lnMachinery	8.027	7.396	8.530	0.366	8.372	6.326	9.539	0.660
minTemp	10.596	8.333	13.143	1.456	6.134	1	12	2.552
invRate	0.010	0.000	0.075	0.017	0.068	-0.044	0.255	0.064
unempRate	0	-0.046	0.068	0.032	0	-0.059	0.063	0.018
unskilledLab	0.200	0.160	0.230	0.023	0.171	0.086	0.270	0.044

Table 6: Descriptive Statistics

	Post-lasso	Std. Error
$\rho$	0.630***	(0.049)
$\ln Land * \rho$	-0.004	(0.007)
$\text{vineSpec} * \rho$	-0.014	(0.022)
$\text{invRate} * \text{unempRate}$	-10.115*	(5.235)
$\text{weather}_{2002}$	0.109***	(0.035)
$\text{france}$	-0.024	(0.063)
$\text{dummy}_{2000}$	-0.073**	(0.031)
$\text{dummy}_{2012}$	0.041	(0.030)
$\text{trend}$	0.003	(0.004)
$\ln Rent$	0.040**	(0.019)
$\ln Depreciation$	0.057	(0.049)
$\ln Taxes$	0.050*	(0.026)
$\ln OtherInputs$	0.011	(0.028)
$\ln MachineryCosts$	0.052*	(0.027)
$\text{otherInputs}^2$	0.053	(0.125)
$\text{otherInputs}^3$	-0.110	(0.130)
$\text{post}$	0.012	(0.037)
$\text{treated}$	0.024	(0.037)
$\delta$	0.089**	(0.037)
constant	-0.552**	(0.273)
F Statistic	482.8***	
Adjusted R <sup>2</sup>	0.922	
Notes:	***p=.01; **p=.05; *p=.1	

Table 7: Post-lasso model estimates

## References

- Abadie, A., Diamond, A., Hainmueller, J., 2010. Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. *Journal of the American Statistical Association* 105, 493–505.
- Abadie, A., Diamond, A., Hainmueller, J., 2011. Synth: An R package for synthetic control methods in comparative case studies. *Journal of Statistical Software* 42 (13), 1–17.
- Abadie, A., Gardeazabal, J., 2003. The economic costs of conflict: A case study of the Basque country. *American Economic Review* 93 (1), 112–132.
- Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., Mitton, T., 2016. The value of con-

- nections in turbulent times: Evidence from the United States. *Journal of Financial Economics* 121 (2), 368–391.
- Achen, C., 2000. Why lagged dependent variables can suppress the explanatory power of other independent variables. Prepared for the Annual Meeting of the Political Methodology Section of the American Political Science Association, UCLA, July 20-22.
- Anderson, T. W., Hsiao, C., 1981. Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76 (375), 598–606.
- ASGI, 2015. Infractions identified by the Italian Association for Juridical Studies on Immigration (ASGI) on sanctions against employers of illegally staying third-country nationals.
- Autor, D., 2003. Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics* 21 (1), 1–42.
- Baccaglio, M., 2016a. Produzione di vino e superfici vitate - aggiornamento 2011. Available at [inumeridelvino.it/2012/07/sicilia-produzione-di-vino-e-superfici-vitate-aggiornamento-2011.html](http://inumeridelvino.it/2012/07/sicilia-produzione-di-vino-e-superfici-vitate-aggiornamento-2011.html).
- Baccaglio, M., 2016b. Produzione doc e docg per denominazione - dati 2014. Available at [inumeridelvino.it/2016/06/produzione-doc-e-docg-per-denominazione-dati-2014.html](http://inumeridelvino.it/2016/06/produzione-doc-e-docg-per-denominazione-dati-2014.html).
- Balkan, B., Tumen, S., 2016. Immigration and prices: Quasi-experimental evidence from Syrian refugees in Turkey. *Journal of Population Economics* 29 (3), 657–686.
- Bandiera, O., 2003. Land Reform, the Market for Protection, and the Origins of the Sicilian Mafia: Theory and Evidence. *The Journal of Law, Economics, and Organization* 19 (1), 218.
- Belloni, A., Chen, D., Chernozhukov, V., Hansen, C., 2012. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80 (6), 2369–2429.
- Belloni, A., Chernozhukov, V., 2013. Least squares after model selection in high-dimensional sparse models. *Bernoulli* 19 (2), 521–547.

- Belloni, A., Chernozhukov, V., Hansen, C., 2013. Inference for high-dimensional sparse econometric models. *Advances in Economics and Econometrics, Tenth World Congress, Vol. 3: Econometrics*, Cambridge University Press: Cambridge, 245–295.
- Borjas, G. J., 2015. The wage impact of the Marielitos: A reappraisal. *National Bureau of Economic Research* 21588.
- Borjas, G. J., 2016. The wage impact of the Marielitos: Additional evidence. *National Bureau of Economic Research* 21850.
- Breusch, T., Pagan, A., 1980. The Lagrange Multiplier Test and its applications to model specification in econometrics. *Review of Economic Studies* 47 (1), 239–253.
- Breusch, T. S., 1978. Testing for autocorrelation in dynamic linear models. *Australian Economic Papers* 17 (31), 334–355.
- Buuren, S., Groothuis-Oudshoorn, K., 2011. mice: Multivariate imputation by chained equations in R. *Journal of statistical software* 45 (3).
- Card, D., 1990. The impact of the Mariel boatlift on the Miami labor market. *ILR Review* 43 (2), 245–257.
- Carrington, W. J., De Lima, P. J., 1996. The impact of 1970s repatriates from Africa on the Portuguese labor market. *ILR Review* 49 (2), 330–347.
- Cavallo, E., Galiani, S., Noy, I., Pantano, J., 2013. Catastrophic natural disasters and economic growth. *The Review of Economics and Statistics* 95 (5), 1549–1561.
- Ceritoglu, E., Yunculer, G., Burcu, H., Torun, H., Tumen, S., 2015. The impact of Syrian refugees on natives’ labor market outcomes in Turkey: Evidence from a quasi-experimental design. *IZA Discussion Paper* 9348.
- Chernozhukov, V., Chetverikov, D., Kato, K., 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics* 41 (6), 2786–2819.
- Chernozhukov, V., Goldman, M., Semenova, V., Taddy, M., Dec. 2017. Orthogonal machine learning for demand estimation: High dimensional causal inference in dynamic panels. *ArXiv e-prints* (arXiv:1712.09988). Available at [adsabs.harvard.edu/abs/2017arXiv171209988C](https://adsabs.harvard.edu/abs/2017arXiv171209988C).

- Chernozhukov, V., Hansen, C., Spindler, M., 2016. hdm: High-dimensional metrics. *R Journal* 8 (2), 185–199.
- Choi, I., 2001. Unit root tests for panel data. *Journal of International Money and Finance* 20 (2), 249–272.
- CMO, 2008. Regulation for Common Market Organization 479/2008. European Common Organization of Agricultural Markets (CMO).
- Corriere delle Migrazioni, 2015. Non CIE piú nessuno (7 march 2015). Available at [corrieredellemigrazioni.it/2015/03/07/non-cie-nessuno/](http://corrieredellemigrazioni.it/2015/03/07/non-cie-nessuno/).
- Croissant, Y., Millo, G., 2008. Panel data econometrics in R: The plm package. *Journal of Statistical Software* 27 (2), 1–43.
- De Arcangelis, G., Di Porto, E., Santoni, G., 2015. Immigration and manufacturing in Italy: evidence from the 2000s. *Economia e Politica Industriale* 42 (2), 163–187.
- Del Carpio, X. V., Wagner, M. C., 2015. The impact of Syrian refugees on the Turkish labor market. *World Bank Policy Research Working Paper* 7402.
- Driscoll, J. C., Kraay, A. C., 1994. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80 (4), 549–560.
- Dublin Regulation, 2013. Regulation no 604/2013 of the European Parliament and of the Council of 26 June 2013 establishing the criteria and mechanisms for determining the Member State responsible for examining an application for international protection lodged in one of the Member States by a third-country national or a stateless person. *Official Journal of the European Union* of 29 June 2013.
- Dustmann, C., Fasani, F., Frattini, T., Minale, L., Schönberg, U., 2016a. On the economics and politics of refugee migration. *CReAM Discussion Paper Series* 1616, Centre for Research and Analysis of Migration, University College London.
- Dustmann, C., Schönberg, U., Stuhler, J., November 2016b. The impact of immigration: Why do studies reach such different results? *Journal of Economic Perspectives* 30 (4), 31–56.
- Eurispes, 2014. *Eurispes media report* 2014.

- European Commission, 2017. Farm accountancy data network. Available at [ec.europa.eu/agriculture/rica/index.cfm](http://ec.europa.eu/agriculture/rica/index.cfm).
- EUROPOL-EMPACT, 2013. Facilitation of illegal migration. EU Policy Cycle - EMPACT (European Multidisciplinary Platform Against Criminal Threats). Available at [europol.europa.eu/crime-areas-and-trends/crime-areas/facilitation-of-illegal-immigration](http://europol.europa.eu/crime-areas-and-trends/crime-areas/facilitation-of-illegal-immigration).
- EUROPOL-EMSC, 2016. Tackling the organised criminal groups profiting from migrant smuggling. European Migrant Smuggling Centre (EMSC). Available at [europol.europa.eu/about-europol/european-migrant-smuggling-centre-emsc](http://europol.europa.eu/about-europol/european-migrant-smuggling-centre-emsc).
- EUROSTAT, 2017. Labor force survey. Tech. rep., Eurostat.
- Flai-Cgil, 2014. Agromafias and Caporalato, Second Report. Osservatorio Placido Rizzotto Flai-Cgil.
- Flai-Cgil, 2016. Agromafias and Caporalato, Third Report. Osservatorio Placido Rizzotto Flai-Cgil.
- Friedberg, R. M., 2001. The impact of mass migration on the Israeli labor market. *The Quarterly Journal of Economics* 116 (4), 1373–1408.
- FRONTEX, 2016. Migratory routes map. European Border and Coast Guard Agency. Available at [frontex.europa.eu/trends-and-routes](http://frontex.europa.eu/trends-and-routes).
- Gaeta, D., Corsinovi, P., 2014. *Economics, Governance, and Politics in the Wine Market: European Union Developments*. Palgrave Macmillan US.
- Gambetta, D., 1993. *The Sicilian mafia - The Business of Private Protection*. Harvard University Press.
- Giangrande, A., 2017. *Caporalato Ipocrisia e Speculazione*. Vol. 165. *L'Italia del Trucco, l'Italia che siamo*.
- Gobillon, L., Magnac, T., 2016. Regional policy evaluation: Interactive fixed effects and synthetic controls. *The Review of Economics and Statistics* 98 (3), 535–551.
- Godfrey, L. G., 1978. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica* 46 (6), 1293–1301.



- Hadri, K., 2000. Testing for stationarity in heterogeneous panel data. *Econometrics Journal* 3 (2), 148–161.
- Heckman, J., 1981. *Heterogeneity and State Dependence*. National Bureau of Economic Research, Inc.
- Honda, J., 1985. Testing the error components model with non-normal disturbances. *Review of Economics Studies* 52, 681–690.
- Hsiao, C., 2014. *Dynamic Models with Variable Intercepts*, 3rd Edition. *Econometric Society Monographs*. Cambridge University Press.
- Human Rights Watch, 2014. Universal periodic review - italy, march 2014. Tech. rep., Human Rights Watch.
- Hunt, J., 1992. The impact of the 1962 repatriates from Algeria on the French labor market. *ILR Review* 45 (3), 556–572.
- Italian Court of Inquiry on Migration, 2015. Scientific police report on migrants’ reception and identification, and on reception facilities for asylum-seekers and expulsion of illegal migrants (audition of 15 september 2015). *Bulletin of Parliamentary Councils and Commissions*, 106. Available at [documenti.camera.it/leg17/resoconti/commissioni/bollettini/html/2015/09/10/69/comunic.htm](http://documenti.camera.it/leg17/resoconti/commissioni/bollettini/html/2015/09/10/69/comunic.htm).
- Italian Interior Ministry, 2016. Asylum statistics 1990-2016. Commissione nazionale per il diritto di asilo.
- Italian Penal Code, 2009. Law n. 94/2009 item 10bis on Dispositions on public security: Entry and illegal stay in the State territory (15 July 2009). *Gazzetta Ufficiale* n. 170 of 24 July 2009.
- Joint ETI, 2015. Decent work in italian agriculture: Counteracting exploitation of migrant workers in tomato production. Project of Ethical Trading Initiative Norway (IEH), Ethical Trading Initiative (ETI) and Danish Ethical Trading Initiative (DIEH).
- Jorgenson, D. W., Griliches, Z., 1967. The explanation of productivity change. *Review of Economic Studies* 34 (3), 249–283.
- Keele, L., Kelly, N. J., 2006. Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political Analysis* 14 (2), 186–205.

- Lamouria, L., Studer, H., Brewer, H., 1963. Improving the productivity of pruning labor in the vineyard. *California Agriculture* 17 (3), 2–3.
- MigrantSicily, 2016. Drawn-out Landing Operations and Facilities in Free-fall: The Failure of the Italian System Creates Violence and Exploitation. *Migrant Sicily Newsletter*, September 2016. Available at [migrantsicily.blogspot.de/2016/10/migrant-sicily-newsletter-september-2016.html](http://migrantsicily.blogspot.de/2016/10/migrant-sicily-newsletter-september-2016.html).
- Moreno-Galbis, E., Tritah, A., 2016. The effects of immigration in frictional labor markets: Theory and empirical evidence from EU countries. *European Economic Review* 84, 76–98.
- Newey, W. K., West, K. D., 1994. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55 (4), 703–708.
- Okkerse, L., 2008. How to measure labour market effects of immigration: A review. *Journal of Economic Surveys* 22 (1), 1–30.
- Palmisano, L., Sagnet, Y., 2016. *Ghetto Italia. I braccianti stranieri tra capolarato e sfruttamento*. Fandango Editore.
- Peri, G., 2012. The effect of immigration on productivity: Evidence from U.S. states. *The Review of Economics and Statistics* 94 (1), 348–358.
- Peri, G., 2016. Immigrants, productivity, and labor markets. *Journal of Economic Perspectives* 30 (4), 3–30.
- Peri, G., Yasenov, V., 2018. The labor market effects of a refugee wave: Synthetic control method meets the Mariel boatlift. *The Journal of Human Resources*. Published online before print January 30, 2018.
- Pesaran, M. H., 2004. General diagnostic tests for cross section dependence in panels. *Econometric Reviews* 34 (6-10), 1089–1117.
- Press Regione Puglia, 2011. Minori migranti non accompagnati: firma di un accordo con Save the Children. Press Release, 16 February 2011. Available at [regione.puglia.it/index.php?page=pressregione&id=9809&opz=display](http://regione.puglia.it/index.php?page=pressregione&id=9809&opz=display).
- Rubin, D., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 5 (5), 688–701.

- Tumen, S., 2016. The economic impact of Syrian refugees on host countries: Quasi-experimental evidence from Turkey. *American Economic Review* 106, 456–60.
- Venturini, A., Villosio, C., 2008. Labour-market assimilation of foreign workers in Italy. *Oxford Review of Economic Policy* 24 (3), 517–541.