The role of museums in bilateral tourist flows: Evidence from Italy.

Nadia Campaniello*[a,b] and Matteo Richiardi †[c,d]

[a]University of Essex, Department of Economics, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom.

[b]IZA - Institute for the Study of Labor, Schaumburg-Lippe-Strasse 5-9, 53113 Bonn, Germany.

[c]University of Turin, Department of Economics, via Po 53, 10124 Torino, Italy.

[d]LABORatorio Revelli and Collegio Carlo Alberto, via Real Collegio 30, 10024 Moncalieri, Torino.

March 8, 2016

Abstract

This paper estimates the causal relationship between the supply of art and tourist flows. To this aim we use aggregate bilateral data on tourist flows and on museums in the twenty Italian regions. To solve the potential endogeneity of the supply of museums we use three different empirical strategies: we control for bilateral macro-area dummies, we compute the degree of selection on unobservables relative to observables which would be necessary to drive the result to zero and, finally, we adopt a 2SLS approach that uses a measure of historical patronage, the number of noble families, as an instrument for the number of museums. We always find strong evidence of a causal relationship between museums and tourist flows.

Keywords: Demand for the art, museums, noble families, cultural tourism, causality.

JEL codes: H23, R12, Z11, D62

*Corresponding author. Email: ncampa@essex.ac.uk

†Email: matteo.richiardi@unito.it

Acknowledgements: Special thanks go to Giovanni Mastrobuoni for his valuable suggestions and constant encouragement. We are grateful to Bruce Weinberg, Francesc Ortega, Orley Ashenfelter, Eugene Smolensky, Mika Kortelainen and Andrea Vindigni for their useful comments. We also thank all the participants at the seminars at the Industrial Relations Section at Princeton University (Princeton, U.S.), at the Department of Economics at Queens College CUNY (New York, U.S), at the Department of Economics at the University of Essex and at the Collegio Carlo Alberto (Moncalieri, Italy). Finally we gratefully acknowledge the comments by the participants to the 69th annual conference of the International Institute of Public Finance (Palermo, Italy), the 5th Applied Economics Workshop (Petralia Sottana, Italy) and the 4th European Workshop on Applied Cultural Economics (Aydin, Turkey).

1 Introduction

A recent article from The Economist (2013a) shows that the number of museums around the world has risen from about 23,000 two decades ago to at least 55,000 now. In 2012, according to the *American Alliance of Museums*, American museums received 850 million visits, that is more than all the big-league sport events and the theme parks combined together. In England more than half of the adult population visited at least a museum or a gallery in 2012, while in Sweden the percentage is close to 67%. Museum-building is also flourishing in developing countries, where governments want to signal that their countries are culturally sophisticated and want their cities to catch up with the great cities of the world. The rise of a large middle class increases the demand for art consumption: China, for example, is investing large sums of money in culture and currently has almost 4,000 museums (thus doubling the number of museums that it had in 2000)Economist (2013b)[1]. In 2011 China opened 386 new museums - more than one per day. To better understand the magnitude of this growth, just think that at the peak of America's recent museum boom (from the mid-1990s to late-2000s), the number of museums constructed a year was only 20-40 (Johnson and Florence, 2012).

Despite such numbers, very little is known about why this is happening and how it is going to influence the economy. The first thing that comes to mind when thinking about potential channels through which museums might affect the economy is tourism. Indeed, tourism represents the main industry and a sizeable portion of total GDP for many countries. According to the World Travel & Tourism Council, worldwide the direct contribution of tourism to total GDP is estimated to be around 3% employing about 100 million workers. Considering its direct, indirect and induced impacts, tourism accounts for 9.3% of global GDP and 1 in 11 jobs.

A significant portion of tourists is believed to travel to visit cultural attractions like museums, churches, etc. (Herrero et al., 2006, Richards et al., 2001), but apart from simple correlations there is little evidence about the importance of culture in generating tourist flows (Blaug, 2001, Bonet, 2003). Moreover, the relationship between cultural supply and tourism might not be as simple as it might seem at first: localities compete to attract "culture-driven tourists" and to restrain their residents from going to other regions by increasing their supply of cultural

[1]Jeffrey Johnson, the founding director of China Megacities Lab at Columbia University (New York City) called this unprecedented museum building boom the "museumification" of China (Johnson and Florence, 2012)

goods. However, if domestic consumers learn about their true preferences through consumption (Levy-Garboua and Montmarquette, 2003) or become addicted to the arts (Becker and Murphy, 1988, Throsby, 1994, McCain, 1979, Brito and Barros, 2005), an increase in local supply may also stimulate the local demand for culture and induce residents to visit other places in search for more cultural goods.

In this paper we use bilateral data on tourist flows across Italian regions to uncover the relationship between tourism and museums.

There are two reasons why Italian data are well suited for identifying and measuring the relationship between the supply of museums and tourist flows. First, due to its historical heritage Italy accumulated an impressive quantity of cultural supply, which is why it is called the "Bel Paese" (in English: "Beautiful Country").[2] Indeed, Italy has the greatest number of UNESCO World Heritage sites in the world (see UNESCO World Heritage Centre webpage). Still, as shown in Figure 1, there is considerable variation in the supply of museums across regions in Italy that can be exploited to estimate its impact on tourism. Second, the largest part of the Italian supply of museums has been accumulated when mass tourism did not even exist, thus reducing concerns about reverse causality. We also control for a large set of obsevables and unobservables (exploring only variations within macro-regions). We show that such historical supply depends on the historical distribution of noble families across the country, and that such distribution can be used to break the potential endogeneity between tourism flows and the supply of art (museums, etc). The main finding is that regions with a larger supply of museums attract more tourists and retain more local cultural consumers from travelling to other regions in search for art.

The paper is organized as follows. In section 2 we present the empirical strategy. In particular, in the subsection 2.2 we discuss the OLS strategy, while, respectively, in the subsections 2.3, 2.4 and 2.5 we present the three different strategies we use to cope with the potential endogeneity: fixed effects, degree of selection on unobservables relative to observables that would explained away our result, and instrumental variable. In section 3 we discuss our results; in section 4 we perform some robustness checks; conclusions are in section 5.

[2]Dante Alighieri and Francesco Petrarca were probably the first ones to use this expression in their poetic works: "del bel paese là dove 'l sì sona" (Dante Alighieri, Inferno, Canto XXXIII, verse 80) and "il bel paese Ch'Appennin parte e 'l mar circonda e l'Alpe" (Francesco Petrarca, Canzoniere, CXLVI, verses 13-14).

2 Empirical analysis

2.1 Road Map

In this section we describe the data and the methodology we use to estimate the effect of museums on tourist flows. Our empirical analysis is based on a gravitational model estimated using OLS for the 20 Italian regions. The dependent variable are the tourist flows from one region (the region of origin) to the other (the region of destination), while the variable of interest is the difference in the number of museums between the region of origin and that of destination. Given that Italy has 20 regions we have a 20 by 20 matrix, that is 400 observations. Since we are not interested in intra-regional tourism, we end up with 380 observations.

As a first preliminary evidence we show raw data and simple correlations. The arrows in Figure 2 represent outgoing per capita regional tourist flows, and their thickness is proportional to the magnitude of such flows (normalized by the population in the region of destination). The shade of grey of each region is related to the number of per capita museums; darker regions have a larger number of museums. Looking at the figure shorter arrows tend to be thicker, indicating that distance plays an important role in the choice of the destination. Furthermore it seems that tourists prefer regions in the north and centre of Italy which display a higher density of museums (darker shades of grey). Figure 3 shows the raw correlation between the outgoing regional tourist flows (log-per capita) and the difference in the availability of museums between the region of destination and that of origin controlling for the population (log-per capita). From this figure it seems that regions with more museums attract more tourists as there is clearly a positive correlation, with the slope equal to 0.29. But in this figure we do not control for other variables, observable and unobservable, that could affect tourism and bias our results. To rule out the possibility that reverse causality or some omitted variables might bias our results we use three different empirical strategies: we control for bilateral macro-area dummies, we calculate the degree of selection on unobservables relative to observables which would be necessary to drive our result to zero and finally we adopt a 2SLS approach using the number of noble families in Italy during the Renaissance as an instrument for the presence of museums.

2.2 OLS strategy

We use aggregate data on tourism inflows and outflows for the twenty Italian regions, complemented with other geographic data and with data on the supply of museums, in order to estimate a model of tourism demand [3]. In particular, we use a gravity model, a spatial model where the degree of interaction between two geographic areas (tourist flows in our case) varies directly with the size of population in the two areas and inversely with the square of the distance between them (Witt and Witt, 1995). To isolate the effect of cultural goods on tourism we control for factors that might be correlated with both the supply of art and tourism, like income, geographical characteristics, etc. Lim (1997) compares all methods used in around 100 published empirical studies of international tourism demand and identifies the most widely used specifications. The dependent variable is generally classified as tourist arrivals and/or departures, tourist expenditures and/or receipts and length of stay, while the explanatory variables are usually income, transportation costs, relative prices, exchange rates and qualitative factors such as destination attractiveness and tourists' attributes (like gender, age, education level and occupation).

We test whether the sum of coefficients of the museums in the region of origin, β_o, and in that of destination, β_d, is equal to zero. In other words, we test whether it is the difference in the availability of museums between regions (M_d-M_o) that really matters. An advantage of using differences as opposed to the two variables taken separately (M_d and M_o) is that by construction differences will vary at the bilateral level. Since we cannot reject that the coefficients sum up to zero, we are going to use the difference in the number of museums in the region of destination and in the region of origin as our variable of interest (see footnote 13).

We use bilateral data on tourism flows and differences in the number of museums between regions in the year 2006. Since Italy has a rather static supply of museums almost the entire variation in the number of museums is across space rather than over time. Moreover, the instrument that we will use later in the 2SLS, based on the historical presence of art patronage (the number of noble families during Renaissance in Italy), is fixed over time as many historical

[3] Despite the universally recognized importance of culture as a source of attraction for tourism, data on cultural tourism are still very limited. Information on the relevance of cultural tourism is scattered and indirect, and often based on *ad hoc* surveys.

instruments are [4].

We use the following specification:

$$\log T_{od} = \beta_{do}(\log M_d - \log M_o) + \beta_o X_o + \beta_d X_d + \beta_\gamma \log Dist_{od} + \mu_{od} \quad (1)$$

where o is the region of origin, d the region of destination. T_{od} is the *per capita tourist flow* from region o (origin) to region d (destination), M_o and M_d are, respectively, indicators of the *supply of (per capita) museums* in the regions of origin and destination [5], X_o and X_d are other characteristics of the two regions (like *income*, opportunity for *mountain* or *sea tourism*, etc.), $Dist_{od}$ is the *distance* between the capital cities in the two regions. The *price of tourism* is generally based on travel cost and on relative prices, that is the difference in the price levels in the regions of origin and destination. We measure travel cost with the *distance* between the capital cities of the regions of origin and destination (Walsh, 1996). To proxy for relative prices across regions we use the *Consumer Price Index*. In order to capture any residual difference in the attractiveness of regions within macro-areas we add *landscape characteristics* (possibility of trekking/hiking/skiing, sea tourism, presence of natural parks). To measure them we use the following variables: *Mountains*, that is the ratio between the mountain area and the total area of a region; *Ski*, that is a dummy equal to 1 if the region hosts ski resorts; *Mountain x Ski* is the interaction between the variables *Mountains* and *Ski*; *Parks* that is the ratio between the surface covered by parks and the total surface of a region; *Coasts* that is the ratio between the coastline length of a region and the total coastal length of Italy. Note that any additional attractiveness is captured by the number of *Foreign tourists* in a region (per capita). The data sources are reported in the Appendix. Table 1 shows the descriptive statistics of the variables and outlines some characteristics of the Italian regions: most of the variables we consider in our analysis vary considerably; income is distributed unevenly, in particular, the South is relatively poor and the North is relatively rich, despite similar levels of education; Italy's dramatic population aging drives the dependency ratio up to almost 57%.

In our specification we cluster the standard errors at both the region of origin and destina-

[4] See for example settler mortality in Acemoglu et al. (2012), the literacy rate at the end of the 19th century and past political institutions in Tabellini (2010) and the presence of a bishop before the year 1000 and foundation by Etruscans in Guiso et al. (2008).

[5] Note that here each museum is treated symmetrically no matter the importance, but that later we will use different sources to check robustness.

tion level (two-way clustering). Cameron and Golotvina (2005) suggest that in cross-sectional regression models for region-pair data, such as gravity models, that allow for the presence of region-specific errors it is important to cluster the standard errors; if not, OLS standard errors are greatly underestimated. Our main focus is on the sign of the coefficient of cultural endowments *(M_d-M_o)* (the difference in the availability of museums in the region of destination and origin) in the gravity model shown in equation 1. Given the log-log specification, the coefficient of the variable representing the cultural endowment can be interpreted as an elasticity. In principle, we should expect a positive coefficient on (M_d-M_o). A null coefficient would signal that art is not a motivation for tourism from o to d, while a positive and significant coefficient would mean that the cultural supply is effective in attracting tourists from other regions.

2.3 The Fixed Effects Estimator

In addition we can exploit the bilateral nature of the data, restricting the variation that is used to identify the coefficient on the difference in the supply of museums. In particular, we generate up to five macro-areas and combine them by origin and destination (for a total of up to 24 bilateral dummies [6]). When adding such fixed effects we only exploit variation within a pair of origin and destination macro-areas. For example, within the Northeast to South group we use only variation across regions of origin that are located in the Northeast (Emilia-Romagna, Friuli-Venezia Giulia, Trentino-Alto Adige and Veneto) and regions of destinations that are located in the South (Abruzzo, Basilicata, Calabria, Campania, Molise and Puglia). The fixed effects would capture any fixed preference for a set of similar region of destination that is common across a set of similar regions of origin (e.g. preferences for climatic, geographic, or cultural differences between the set of regions). In order to capture any residual variation that might bias the coefficients on the supply of museums we control for several other variables that are likely to influence tourism flows as well as museums (for both, origin and destination regions): resident population, per capita income, as well as the Gini coefficient, education, and the demographic dependency ratio. [7]

[6] There are $5^2 = 25$ combination available and we drop one dummy variable from the regressions.

[7] The *population* of the region of origin represents the potential demand for tourism. The population of the region of destination is likely to influence its attractiveness as well, at least through visits to friends and relatives. The budget constraint of tourists depends on the level of income in the region of origin (thus we control for the *per capita regional income*) and possibly also on its distribution as measured by *the regional Gini index*. We also include two other socio-demographic variables of the region of origin in the model: the *level of education*,

2.4 Degree of selection on unobservales relative to observables

Even though we control for many observables that are likely to be correlated with both the number of museums and tourist flows, our results might still be biased by unobservable factors that vary within macro-areas. To rule out the possibility that omitted variables might bias our results we compute the degree of selection on unobservables relative to observables (the so called "implied ratio") which would be necessary to drive the result to zero. This approach is based on the idea that the bias generated by the observed controls provides information on the bias that is generated by the unobserved ones (Altonji et al., 2005, Oster, 2013). In other words we investigate how the inclusion of additional regressors change the coefficient on our variable of interest *(M_d-M_o)*. If the coefficient on the difference in the number of museums change substantially it would be possible that the inclusion of other regressors would significantly reduce the estimated effect. On the contrary, if the coefficient does not vary substantially we are more confident of the causal interpretation of the relationship. [8]

2.5 Instrumental variable strategy

As an alternative to the degree of selection strategy we devise an instrument that is plausibly exogenous: the number of Italian noble families from a region as an instrument for museums. There is an historical explanation for why this is likely to be a valid instrument. Between the XV and the XVIII century Renaissance characterized Europe and in particular, Italy, that was well known for its cultural achievements. Art was often financed by wealthy noble families and important representatives of the Church (high ranking officers such as the Pope, cardinals, and bishops) who used patronage of the arts to signal their status, power and, for religious commissions, piety (Nelson and Zeckhauser, 2008), and not as a mean to attract tourism.

Wealth inequality was an important driver of the Renaissance. Artistic developments depended on the patronage of an elite of very wealthy people who wanted to distinguish themselves

---

measured by the percentage of people with at least a middle school diploma, and the *demographic dependency ratio*, equal to the ratio between the population aged 65 or over and the population aged 20-64. The level of education is expected to be positively correlated with tourism, while the demographic dependency ratio has an a priori ambiguous effect on tourist flows (traveling for business being more likely for prime age individuals, while pilgrimages being more frequently associated with the elderly).

[8]These bounds are now often computed in empirical work. For example this approach has been used by Bellows and Miguel (2009) in their study on the impact of the Sierra Leone civil war on individuals who have been victimised in terms of their postwar socio-economic status, their political mobilization and engagement, by Nunn and Wantchekon (2011) in their paper on the impact of slave trade on mistrust in Africa and by Adhvaryu et al. (2014) in their paper on the effect of cocoa price shocks at birth on adult mental health outcomes.

from those of lesser status and they needed to demonstrate “magnificence” (Hollingsworth, 1994): to be rich meant to be a patron of the arts (Pullan, 1973, Goldthwaite and Gerulaitis, 1995).

Many of the most important and visited Italian museums were built before the start of mass tourism. Only the rise of the bourgeoisie in the XIX century caused the move from patronage to a publicly supported system of the arts, a system where investments could depend on tourism flows. In particular, tourism began in the XVIII and XIX centuries, when European aristocrats and rich bourgeois started to travel to Mediterranean countries for the so called “Grand Tour” (Towner and Wall, 1991). This elitarian form of tourism was replaced by mass tourism in Western Europe only after World War II (Costa, 1989). Hence cultural goods dating back more than 70 years from now were not created as a response to (high or low) tourist flows; they were just a way to celebrate power ad magnificence of the patrons. Some famous examples are the “Vatican Museums” in Rome, the “Galleria degli Uffizi” (Uffizi Gallery) in Florence, the “Palazzo Ducale” (Doge’s Palace) in Venice, the “Reggia di Caserta” (the Royal palace of Caserta) in the Kingdom of Naples, or the “Reggia di Venaria Reale” (the Royal palace of Venaria Reale) in the Duchy of Savoy.

Looking at the general ranking of the most visited Italian museums in 2011 (Il Giornale dell’Arte.com, May 2012, see Table 2), the mentioned museums are ranked, respectively: first (with 5,078,004 visitors), second (with 1,766,345 visitors), third (with 1,403,524 visitors), tenth (with 571,368 visitors) and eleventh (with 534,777 visitors).

The Vatican Museums (included in the Lazio region in our dataset) were founded in the XVI century by Pope Iulius II, as a part of a more general project aimed at making Rome an impressive centre that could demonstrate the prestige of the Pope as the supreme head of the church patronage.

The Uffizi Gallery is, nowadays, the most important and visited museum in Florence. The building of the Uffizi palace started in 1560 when Cosimo de’ Medici, first Grand Duke of Tuscany, was consolidating his power, with the aim to host the administrative and judicial offices. He clearly filled the palace with art to impress those who visited the palace and to show his economic and political power.

The Doge’s Palace in Venice (the Palace of the head of state, the “Doge”) was the headquarter

of power of the Venetian Republic, hosting the political institutions of the state. It is regarded as a masterpiece of Gothic architecture. It acquired its actual aspect in the Renaissance period, when famous architects and painters worked on it.
The Royal Palace of Caserta was started in 1752 for Charles III of Naples as the new centre of the Kingdom of Naples and it is a masterpiece of the baroque architecture. Since 1997 it is a UNESCO World Heritage Site.
The Royal Palace of Venaria Reale was one of the royal residences of Savoy located in Venaria Reale, close to Torino, in northern Italy. The construction of the palace started in 1675 under the patronage of the Duke Carlo Emanuele II, who wanted to celebrate his magnificence building a hunting residence that could compete with the Palace of Versailles In France.

To collect data on patrons in the Renaissance we went as far back in time as possible through the story and genealogy of the around 1,800 noble families in Italy in the "The Golden Book of Italian Nobility" (*Libro d'oro della Nobiltà Italiana*). Such publication has a comprehensive list of the Italian noble families with the indication of their origins, which predates mass tourism. The process of expropriation of important buildings owned by nobel families started with the unification of Italy (1861), continued in the 1920s and 30s by the Mussolini government, but gained real momentum after World War II. In 1946 the Italian Savoy Kingdom was replaced by a Republic and titles of nobility lost their legal status. With the Republican Constitution all property owned by the Savoy family was transferred to the State (e.g. the Royal Palace of Venaria Reale, the Royal Palace of Turin, etc.). But the State expropriated many additional buildings owned by other families, as for example the Villa Doria Pamphilj in 1957, and Palazzo Barberini in 1949.

Moreover, in 1950 the Italian government expropriated land from large-scale land properties, called latifundia, which were mainly in the hands of noble families. The sudden loss of agricultural revenues forced many families to give up their real estate properties.

The data we collected include records on high ranking officers of the Church, which most times were second-born sons of noble families. Amidst the 28 Popes who were heading the Church between the beginning of the XV and the end of the XVII century, 24 belonged to noble families (restricting our attention to the 24 Italian Popes, 21 were of noble origins).
Despite the fact that many of these buildings became museums before the advent of mass

tourism the origin of nobel families might proxy for additional amenities, like wealth, income, landscape, etc. For this reason it is important to control for these amenities, meaning that the IV is only conditionally independent. Another objection could be that noblemen are a subset of tourists thus violating the exclusion restriction. But the number of noble families is extremely small compared to the size of tourist flows, and the region of origin of the noble families is in most cases different from the region where they reside today.

Table 3 shows the number of noble families in each Italian region. There is substantial variability across regions and most of the museums are located in the Central and Northern part of the country. In Figure 4 we plot the difference in the presence of noble families in the region of destination and in the region of origin (over population) and the difference in the presence of museums in the region of destination and in the region of origin (over population) at the regional level. The correlation between noble families (per capita) and museums (per capita) is strongly positive. Below we show that the correlation survives even in the 2SLS setup, after controlling for other regressors, including the amenities.

3 Results

Table 4 shows the coefficients of the gravity model estimated by OLS (table 8 in the Appendix shows the results of the OLS with all the regressors we use in our specification). We use both robust standard errors (in the left parenthesis) and clustered standard errors at the region of origin and destination (in the right parenthesis). In the first column we do not control for bilateral macro-area dummies, while in the second column we control for 3 bilateral macro-area dummies [9], in the third for 8 bilateral macro-area dummies [10] and in the fifth for 24 bilateral macro-area dummies [11]

[9]We generated two area dummies: *North* that includes the region of Liguria, Lombardia, Piemonte, Valle d'Aosta, Emilia-Romagna, Friuli-Venezia Giulia, Trentino-Alto Adige, Veneto, Lazio, Marche, Toscana, and Umbria and *South* that includes the region of Abruzzo, Basilicata, Calabria, Campania, Molise, Puglia, Sardegna and Sicilia.

[10]We generated three area dummies: *North* that includes the region of Liguria, Lombardia, Piemonte, Valle d'Aosta, Emilia-Romagna, Friuli-Venezia Giulia, Trentino-Alto Adige and Veneto, *Center* that includes the region of Lazio, Marche, Toscana, and Umbria and *South* that includes the region of Abruzzo, Basilicata, Calabria, Campania, Molise, Puglia, Sardegna and Sicilia.

[11]We generated five area dummies: *Northwestern* that includes the region of Liguria, Lombardia, Piemonte, Valle d'Aosta, *Northeastern* that includes the region of Emilia-Romagna, Friuli-Venezia Giulia, Trentino-Alto Adige and Veneto, *Central* that includes the region of Lazio, Marche, Toscana, and Umbria, *South* that include the region of Abruzzo, Basilicata, Calabria, Campania, Molise and Puglia, *Islands* that include the region of Sardegna and Sicilia. There are $5^2 = 25$ combination available and we drop one dummy variable from the

When adding a larger number of bilateral macro-area dummies we are restricting the available variation in the data, controlling for an increasing set of unobserved fixed preferences across macro-regions that might bias our coefficient on the log difference in museums (per capita). Not controlling for area dummies the elasticity of the difference in the number of museums in the region of destination and in that of origin is statistically significant and is equal to 0.383. When we add bilateral macro-region dummies we get larger elasticities, and the elasticities get larger as we increase the number of macro-regions (1.469 controlling for 3 bilateral macro-area dummies; it increases to 1.473 controlling for 8 bilateral macro-area dummies and to 1.829 controlling for 24 bilateral macro-area dummies) [12]. This suggests that restricting the variability tends to reduce a bias that is driving the coefficients towards 0. This is consistent with local governments with disappointingly low numbers of visitors opening up a larger number of museums, or, simply, with attractive regions having no interest in managing public museums. Controlling for bilateral macro-area fixed effects the coefficient on the museums variable increases dramatically meaning that there are some important unobserved preferences that affect bilateral tourism within bilateral macro-regions (e.g. over the last 50 years Italy has experienced large-scale migration flows from the South which is poorer and has fewer museums to the North of the country which is richer and has more museums. Most of these internal migrants have maintained strong links with their region of origins where they still have relatives. Part of the flows we observe might be driven by these migrants, and more generally by individuals that are attracted to the south despite the smaller number of museums. The bilateral macro-region effects would be able to capture the phenomena, reducing the bias of the estimates. We cannot observe this kind of tourism but it is likely to be quite large)[13]. In the last 2 row of table 4 we compute the implied ratios and the selection on the unobservables that would be needed to drive our results to zero.

regressions.

[12]We also run the regressions using a Poisson estimator, as suggested by Silva and Tenreyro (2006): under heteroskedasticity, the parameters of log-linearized models estimated by OLS might lead to biased estimates of the true elasticities. The estimated effect of the difference in the number of museums is positive and significant at 1% level (the coefficient on M_d-M_o is equal to around 0.29 without bilateral fixed effects and increases up to 0.89 with bilateral fixed effects.

[13]Our preferred specification is the one that uses the largest number of bilateral area dummies. The specification in first differences between destination and origin that we use relies on the assumption that adding a museum in the region of destination has the same effect as reducing the number of the museum in the region of origin. For this reason we also regressed tourist flows on the number of museums in destination and in origin separately and then test the assumption that the coefficients sum up to zero or, in another words, are symmetric. We find that the two coefficients taken separately are not significantly different from zero (the p-value is equal to 0.21 with robust standard errors and to 0.13 with clustered standard errors).

In all the specifications we find ratios far below 1 meaning that, in fact, the coefficients are even larger. Without bilateral macro-area dummies the selection on unobservables would have to be almost 8 times as strong as selection on the observables to produce a treatment effect of zero and should go in the opposite direction because its sign is negative. When we use bilateral macro-area dummies we find that the selection on the unobservables would have to be between 2.58 and 4.05 to explain away the full estimated effect and should go in the opposite direction because its sign is negative. Using the heuristic cutoff equal to 1 suggested by Altonji et al. (2005) and Oster (2013) for the ratio between selection on observables and selection on the unobservables (meaning that the selection of the observable is identical to the one on the unobservables), the coefficient on the variable of interest would actually be even larger (43% without bilateral macro-area dummies and 200-228% with bilateral macro-area dummies)[14]. These results imply that it is highly unlikely that our estimates can be fully attributed to unobserved heterogeneity.

Let us discuss the size of the effects that we estimate. If we take a region with 200 museums, which is close to the average number (238 museums) and we open additional 20 museums, the expected number of incoming tourists would increase by about 3.383% (10% $\times$ 0.383) when using our most conservative OLS estimates. Assuming a close-to-average annual flow of 100,000 visitors from each of the other 19 regions, this amounts to 64,277 more visits inside the region [15].

We now turn to the IV estimates. The results from the first stage, the reduced form and the IV (2SLS) regression are shown in Table 5. The coefficient on the number of noble families is positive and significant, equal to 0.318. Since none of the regressors in the first stage vary at the bilateral level the reported coefficients are all symmetric. We use both robust and two-way cluster-robust standard errors by region of origin and region of destination. The first stage F-statistic of the excluded instrument is equal to 943.33 using robust standard errors and to 144.11 using two-way cluster-robust standard errors, that is well above the rule of thumb of 10 indicated in the literature on weak instruments (Bound et al., 1995, Stock and Yogo, 2002). Column 2 shows the estimates for the reduced form. The coefficient on the number of noble

[14] One reason to favor this cutoff is that researchers typically focus their data collection efforts (or their choice of regression controls) on the controls they believe ex ante are the most important (Angrist and Pischke, 2010)

[15] To this, we should add the increase in the number of foreign visitors

families is positive and significant when we use robust standard errors (it is almost significant, at 14%, when we cluster the standard errors) and equal to 0.073. The last column in Table 5 reports the results of the IV (2SLS). The coefficient M_d-M_o is equal to 0.229 an its is close to that of the OLS estimation without bilateral area dummies. These results confirm that museums help attracting tourists from other regions and retaining the local residents to go to other regions to consume art [16]. When we introduce bilateral area fixed effects in the 2SLS regression the first stage F-statistic is far below the rule of thumb of 10 (2.47 with 2 bilateral area dummies, 2.80 with 8 bilateral area dummies and 4.51 with 24 bilateral area dummies) indicating that the instrument is too weak. The regression of the number of noble families on just the bilateral area fixed effects has a R-squared that is around 0.5 meaning that fixed-effects explain most of the variation. For this reason we cannot use bilateral area fixed effects in the IV specification.

4 Robustness checks

We perform different robustness checks (see tables 6 and 7) to make sure that our results do not depend on the particular specification we used. Like we did in the main regressions we use both robust standard errors and two-way cluster-robust standard errors by region of origin and region of destination. We use four different specifications: the first one (column 1) without bilateral macro-area dummies and the other three with, respectively, 3, 8 and 24 bilateral macro-area dummies (column 2-4). Since the OLS estimates appear to be a conservative estimate of the effect of museums on tourist flows, the robustness checks are based on the OLS specifications.

To be sure that our results are not biased by the different dimension of the regions we estimate a weighted regression, weighting for population in the region of origin. Again, the coefficient on *(M_d-M_o)* is significant and positive (its elasticities is between 0.461 without bilateral macro-area dummies and 1.732 with 24 bilateral macro-area dummies).

We estimate a regression without per capita values controlling for the population in the region of origin and in the region of destination. The coefficient on *(M_d-M_o)* is still positive and significant in all the specifications but the first one without bilateral fixed effects (its elasticities

[16] While without bilateral macroareas a Hausman test rejects the hypothesis that there is endogeneity, the instrument varies too little within macroareas to run the IV using such dummies.

is between 0.145 without bilateral macro-area dummies and 0.715 with 9 bilateral macro-area dummies).

We also adopt a specification that includes the fraction of international flight passengers in the region of origin and destination as a proxy for efficient transports: the coefficient on $(M_d\text{-}M_o)$ is still positive and significant (its elasticities is between 0.733 with 9 bilateral macro-area dummies and 2.623 with 24 bilateral macro-area dummies). We consider the number of international passengers because the number of Italian passengers would clearly be endogenous.

In Table 7 we cope with the potential measurement error using two different measures of museums and we also take into account the fact that museums are not the only typology of cultural goods considering other two additional important cultural goods: theater performances and concerts.

First, we take into account as an alternative measure of the number of museums provided by the website "museionline.it", a partnership between Microsoft and Adnkronos Culture, a news agency which collects and constantly updates information on over 3,500 museums in Italy. The coefficient on $(M_d\text{-}M_o)$ is statistically significant. Its elasticity is between 0.282 without bilateral macro-area dummies and 0.539 with bilateral macro-area dummies. Then we use a measure of the (perceived) quality of the museums: the list of the top cultural attractions on the website "tripadvisor.com" at a regional level. The coefficient on $(M_d\text{-}M_o)$ is between 0.237 (without bilateral macro-area dummies) and 0.473 (with 24 bilateral macro-area dummies).

Finally, we perform a robustness check using a composite index (*the cultural index*), that is an aggregated measure of three different cultural goods: museums, theater performances and concerts. The index is constructed with a factor analysis and represents a weighted average of the three cultural measures, where the weights are based on the correlation structure of these variables. The difference in the supply of art between the region of destination and that of origin measured by the cultural index has a positive and significant effect on tourist flows and its elasticity is between 0.260 (without bilateral macro-area dummies) and 0.371 (with 24 bilateral macro-area dummies).

5 Conclusions

To the best of our knowledge, this is the first paper that identifies a causal relationship between the number of museums and tourist flows. Cultural attractions are shown to have a significant effect on tourist flows.

To address the potential endogeneity problem we use a series of different identification strategies: i) a "within" bilateral macro-areas estimator, ii) we exploit the information on the observable variables to infer selection on the unobservable variables adopting Altonji et al. (2005)'s approach and its development by Oster (2013) and, finally, iii) we adopt a Two Stage Least Squares (2SLS) approach that uses the number of Italian noble families who were originally residing in a region as an instrument for the provision of museums. The results are consistent across all methods.

Our instrumental variable for the number of museums, the number of noble families during Renaissance in Italy, could be used for all those countries that experienced art patronage when cultural tourism did not exist. Since art patronage tended to arise wherever a royal or imperial system dominated a society, our instrument could be appropriate for those countries that were ruled by an aristocracy before the XIX century: among others France, Germany, United Kingdom, Spain, the Netherlands, Denmark, Sweden, Belgium and Austria.

Let us conclude saying that ideally one would like to perform a cost-benefit analysis of running a museum. To this aim we would need information not only on the benefits but also on the costs. Regarding the benefits we would need to take into account not only the number of visitors but also other sort of spending that the city could benefit from (including externalities). Regarding the costs we would need balance sheet data for a representative number of museums. Unfortunately these data are not available for Italy.

References

D. Acemoglu, S. Johnson, and J. A Robinson. The colonial origins of comparative development: An empirical investigation: Reply. *The American Economic Review*, 102(6):3077–3110, 2012.

A. Adhvaryu, J. Fenske, and A. Nyshadham. Early life circumstance and mental health in ghana. Technical report, Centre for the Study of African Economies, University of Oxford, 2014.

Joseph G Altonji, Todd E Elder, and Christopher R Taber. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1), 2005.

Joshua Angrist and Jörn-Steffen Pischke. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. Technical report, National Bureau of Economic Research, 2010.

G. S. Becker and K. M. Murphy. A theory of rational addiction. *Journal of Political Economy*, 96(4):675–700, 1988.

J. Bellows and E. Miguel. War and local collective action in sierra leone. *Journal of Public Economics*, 93(11):1144–1157, 2009.

M. Blaug. Where are we now on cultural economics. *Journal of Economic Surveys*, 15(2): 123–143, 2001.

L. Bonet. Cultural tourism. In R. Towse, editor, *A handbook of cultural economics*. Edward Elgar Publishing, 2003.

J. Bound, D.A. Jaeger, and R.M. Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association*, pages 443–450, 1995.

P. Brito and C. Barros. Learning-by-consuming and the dynamics of the demand and prices of cultural goods. *Journal of Cultural Economics*, 29(2):83–106, 2005.

A. Cameron and N. Golotvina. Estimation of models for country-pair data controlling for clustered errors: with applications. *UC Davis, Manuscript*, 2005.

Collegio Araldico. *Libro d'Oro della Nobiltà Italiana*. Number vol. XXV - XXVI in XXII edition. Collegio Araldico, 2004.

N. Costa. *Sociologia del turismo*. Cooperativa libraria IULM, Milan, 1989.

The Economist. Temples of delight. Special report, December 2013a.

The Economist. Mad about museumst. Special report, December 2013b.

R.A. Goldthwaite and L.V. Gerulaitis. Wealth and the demand for art in italy, 1300–1600. *History: Reviews of New Books*, 23(2):77–78, 1995.

L. Guiso, P. Sapienza, and L. Zingales. Long term persistence. Technical report, National Bureau of Economic Research, 2008.

L.C. Herrero, J.Á. Sanz, M. Devesa, A. Bedate, and M.J. Del Barrio. The economic impact of cultural events a case-study of salamanca 2002, european capital of culture. *European Urban and Regional Studies*, 13(1):41–57, 2006.

M. Hollingsworth. *Patronage in Renaissance Italy: from 1400 to the early sixteenth century*. John Murray, 1994.

Jeffrey Johnson and Zoe Alexandra Florence. The museumification of china. Technical report, M+ Matter, 2012.

L. Levy-Garboua and C. Montmarquette. Demand. In R. Towse, editor, *A handbook of cultural economics*. Edward Elgar Publishing, 2003.

C. Lim. Review of international tourism demand models. *Annals of Tourism Research*, 24(4): 835–849, 1997.

R.A. McCain. Reflections on the cultivation of taste. *Journal of Cultural Economics*, 3(1): 30–52, 1979.

J.K. Nelson and R. Zeckhauser. *The patron's payoff: conspicuous commissions in Italian Renaissance art*. Princeton University Press, 2008.

Nathan Nunn and Leonard Wantchekon. The slave trade and the origins of mistrust in africa. *American Economic Review*, 101:3221–3252, 2011.

Emily Oster. Unobservable selection and coefficient stability: Theory and validation. Technical report, National Bureau of Economic Research, 2013.

B.S. Pullan. *A history of early Renaissance Italy: from mid-thirteenth to the mid-fifteenth century*. Allen Lane, 1973.

G. Richards et al. The development of cultural tourism in europe. *Cultural attractions and European tourism*, pages 1–18, 2001.

JMC Santos Silva and Silvana Tenreyro. The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658, 2006.

J.H. Stock and M. Yogo. Testing for weak instruments in linear IV regression. Technical report, National Bureau of Economic Research, 2002.

G. Tabellini. Culture and institutions: economic development in the regions of europe. *Journal of the European Economic Association*, 8(4):677–716, 2010.

D. Throsby. The production and consumption of the arts: a view of cultural economics. *Journal of economic literature*, 32(1):1–29, 1994.

J. Towner and G. Wall. History and tourism. *Annals of Tourism Research*, 18(1):71–84, 1991.

M. Walsh. Demand analysis in irish tourism. *Statistical and Social Inquiry of Ireland*, 1996.

S.F. Witt and C.A. Witt. Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting*, 11(3):447–475, 1995.

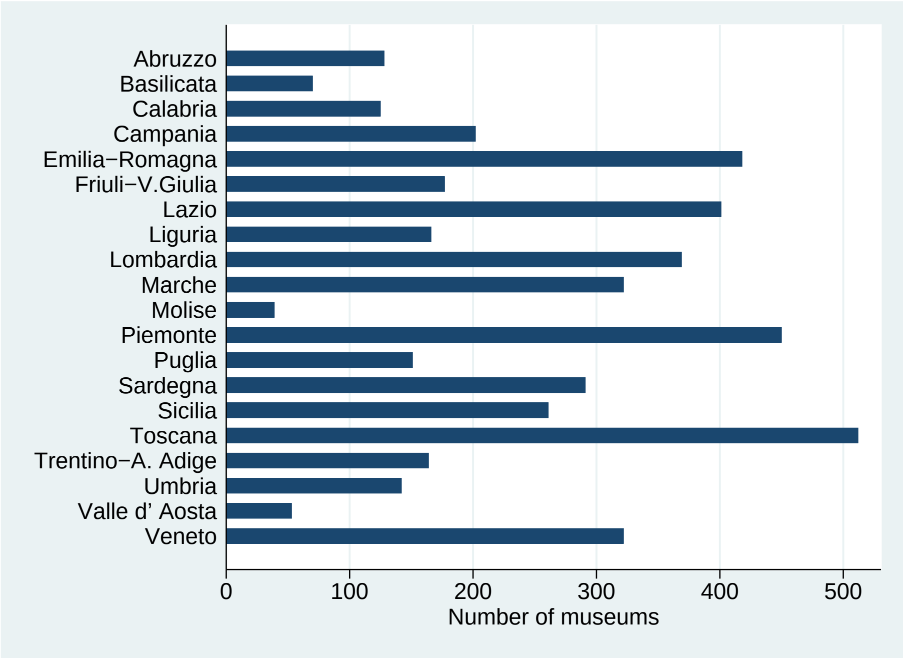

Figure 1: Number of museums by region.

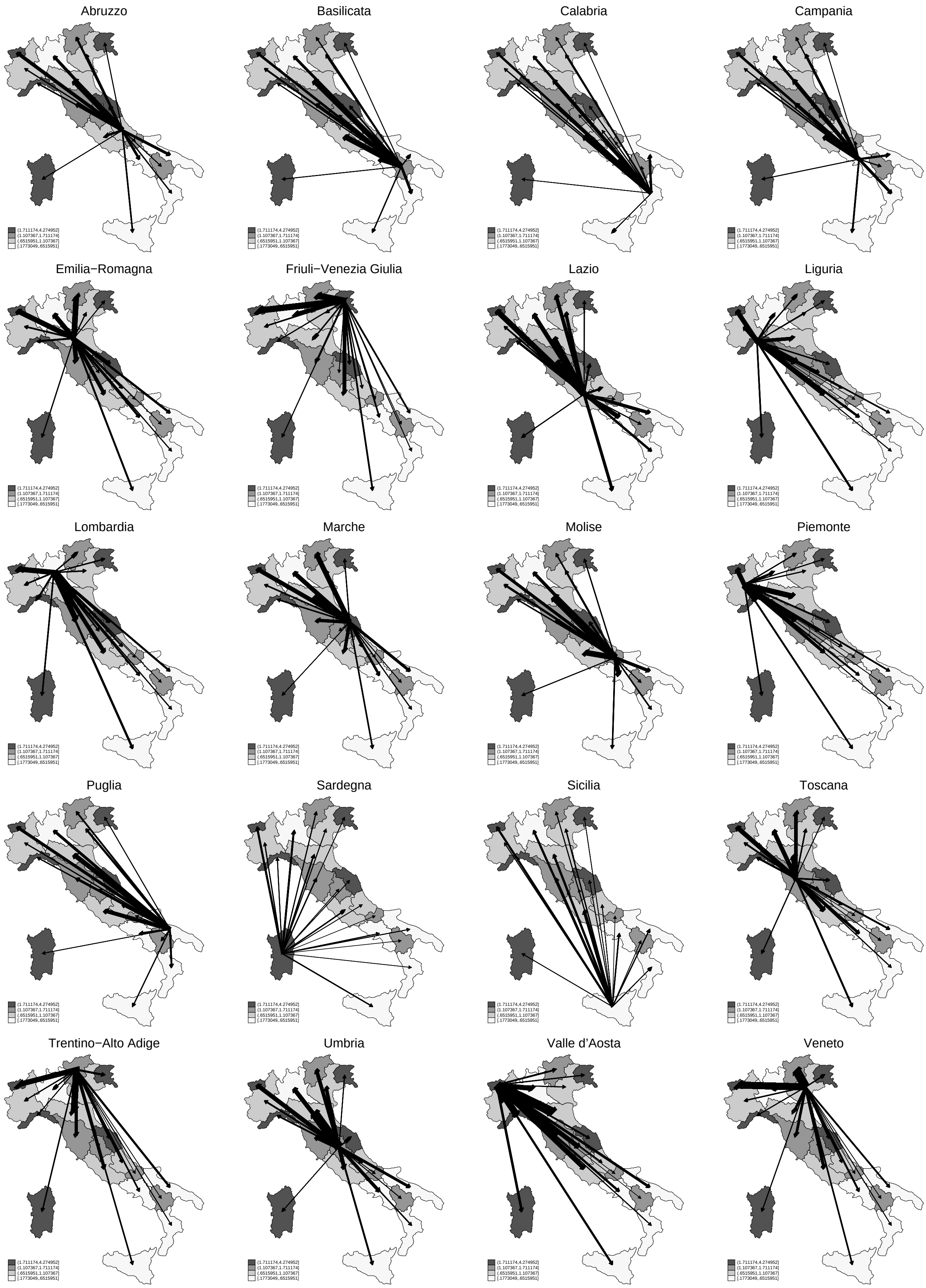

Figure 2: Outgoing regional tourist flows

Notes: The thickness of each line is proportional to the magnitude of the tourist flows normalized by the population in the region of destination. Regions are colored according to the number of museums per capita, darker regions having more museums.

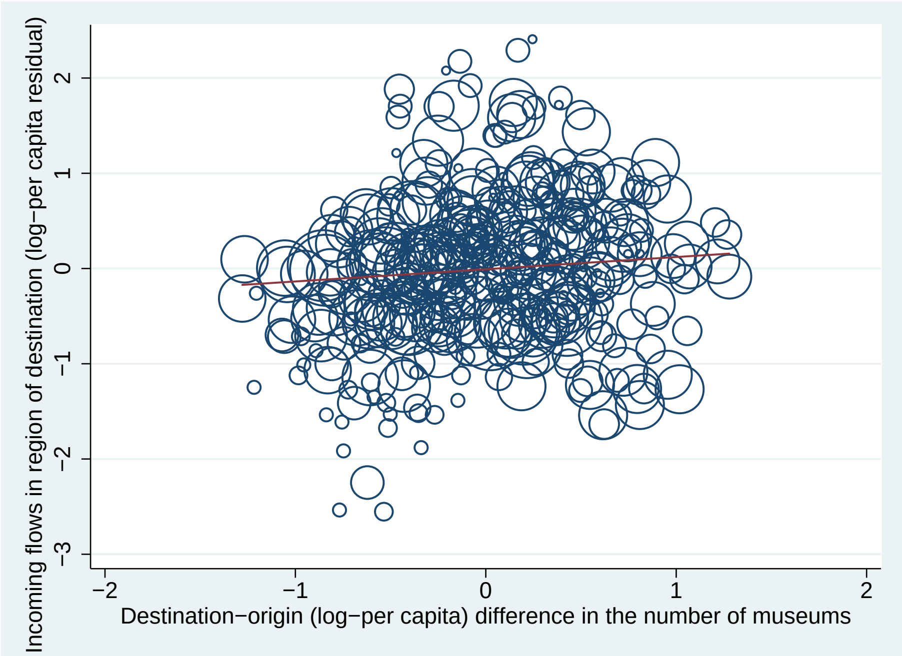

Figure 3: Outgoing tourist flows (per capita) and the difference in the availability of museums between the region of destination and that of origin. We control for the population. Circles are proportional to population size.

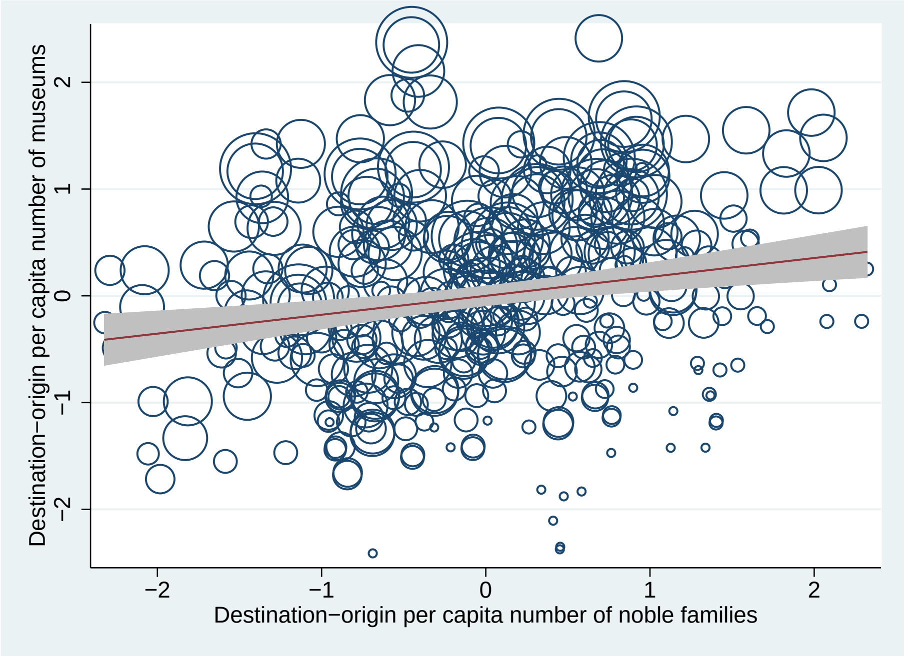

Figure 4: Correlation between the difference in the number of per-capita noble families between the region of destination and that of region (per 100,000 inhabitants) and the difference in the number of per-capita museums between the region of destination and that of region (per 100,000 inhabitants). Circles are proportional to population size.

Table 1: Summary statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Between regions tourist flows	380	107,520	171,134	91	1,464,579
Museums (ISTAT)	380	237	134	42	526
Museums (museionline.it)	380	160	103	18	348
Museums (tripadvisor.com)	380	59	42	3	130
Theatrical performances	380	8,424	748,228	201	27,342
Concerts	380	1,731	1,651	75	6,616
Noble families	380	88	71	2	240
Population (000)	380	2,926	2,353	124	9,475
Regional income (billions Euros)	380	74.2	71.0	4.1	307.7
Distance (km)	380	599	340	105	1,642
Mountain	380	0.42	0.25	0.01	1
Ski	380	0.15	0.36	0	1
Park	380	0.11	0.07	0.02	0.28
Coast	380	0.05	0.07	0	0.26
Secondary education or above	380	0.73	0.03	0.69	0.80
Foreign Tourists	380	17,137.7	15,632.79	779	50,309
CPI	380	100.4	7.3	88.0	113.3
Gini Index	380	0.29	0.02	0.26	0.33
Dependency Ratio	380	50.2	3.3	42.8	56.7
International flight passengers	380	0.05	0.10	0	0.37

Regional income (in Euro) is divided by 1,000,000,000; population by 10,000, *Tourist flows* by 1000, *Foreign Tourists* by 1000 and *distance* (km) by 100.

Table 2: Italian museums by number of visits.

Ranking	Museum	Region	Visitors	Century
1	Musei Vaticani	Lazio	5,078,004	XVI
2	Galleria degli Uffizi	Toscana	1,766,345	XVI
3	Palazzo Ducale	Veneto	1,403,524	XIV
4	Galleria dell'Accademia	Toscana	1,252,822	XVIII
5	Museo Nazionale di Castel Sant'Angelo	Lazio	981,821	XIII
6	Museo Centrale del Risorgimento	Lazio	821,000	XIX-XX
7	Museo Argenti, Museo Porcellane, Boboli	Toscana	714,224	XV
8	Museo Nazionale del Cinema	Piemonte	608,448	XIX
9	Museo delle Antichità Egizie	Piemonte	577,042	XVII
10	Reggia di Caserta	Campania	571,368	XVIII
11	Reggia di Venaria Reale	Piemonte	534,777	XVIII
12	Museo di Palazzo Vecchio	Toscana	533,218	XII-XIV
13	Museo del Novecento	Lombardia	522,100	XX
14	Museo e Galleria Borghese	Lazio	506,368	XVII
15	Musei Capitolini	Lazio	469,351	XVIII

Table 3: Noble families.

Region	**Noble families**
Abruzzo	17
Basilicata	7
Calabria	52
Campania	147
Emilia-Romagna	145
Friuli-V.Giulia	39
Lazio	120
Liguria	99
Lombardia	240
Marche	90
Molise	2
Piemonte	216
Puglia	33
Sardegna	27
Sicilia	122
Toscana	183
Trentino-A. Adige	27
Umbria	55
Valle d' Aosta	2
Veneto	137

Table 4: Estimates of the OLS regressions.

	Log Tourist flows_od (per capita)			
	(1)	(2)	(3)	(4)
Log Museums_d (per capita) - Log Museums_o (per capita)	0.383	1.469	1.473	1.829
	(0.086) (0.108)	(0.219) (0.088)	(0.219) (0.094)	(0.259) (0.195)
Log Population_o	-0.099	-0.802	-0.790	-0.772
	(0.057) (0.080)	(0.138) (0.083)	(0.138) (0.105)	(0.139) (0.117)
Log Population_d	0.863	1.539	1.568	1.834
	(0.060) (0.124)	(0.141) (0.068)	(0.140) (0.069)	(0.142) (0.130)
Log Distance	-0.654	-0.723	-0.701	-0.734
	(0.054) (0.092)	(0.064) (0.096)	(0.079) (0.115)	(0.057) (0.097)
Log Regional Income_o (per capita)	0.352	2.491	2.494	2.674
	(0.399) (0.454)	(0.572) (0.243)	(0.566) (0.237)	(0.576) (0.224)
Log Regional Income_d (per capita)	-2.659	-5.131	-5.166	-6.123
	(0.410) (0.716)	(0.571) (0.434)	(0.549) (0.232)	(0.548) (0.350)
Log Education_o	2.272	-1.422	-1.751	3.522
	(1.004) (0.580)	(1.242) (0.548)	(1.335) (0.674)	(1.582) (0.550)
Log Education_d	-5.535	-2.639	-3.311	-1.225
	(1.094) (2.107)	(1.260) (0.851)	(1.297) (1.009)	(1.565) (1.476)
Log Foreign Tourists_o (per capita)	0.094	-0.545	-0.520	-1.141
	(0.073) (0.091)	(0.124) (0.078)	(0.124) (0.096)	(0.230) (0.116)
Log Foreign Tourists_d (per capita)	0.950	1.525	1.580	1.672
	(0.069) (0.110)	(0.126) (0.077)	(0.125) (0.065)	(0.221) (0.167)
Number of bilateral area dummies	0	3	8	24
Observations	380	380	380	380
R-squared	0.874	0.886	0.889	0.921
Selection on the unobservables that would drive our results to zero	-7.652	-2.588	-2.685	-4.055
Coefficient on the variable of interest with a cutoff equal to 1	0.433	2.037	2.021	2.28

Regression results for the log of region-to-region tourist flows (divided by the population in the region of origin) *Log Tourist flows_od (per capita)* on the log difference in the number of museums in the region of destination and origin (per capita) M_d-M_o with all the regressors we use in our specification. For the complete list of the regressors we use in our specification see Table 8 Column 1 shows results not controlling for bilateral area dummies. Column 2 controls for 3 bilateral area dummies (north-north, north-south and south-north). Column 3 controls for 8 bilateral area dummies (north-north, north-south, north-center, center-north, center-center, center-south, south-north, south-center). Column 4 controls for 24 bilateral area dummies (northwest-northwest, northwest-northeast, northwest-center, northwest-south, northwest-islands, northeast-northwest, northeast-northeast, northeast-center, northeast-south, northeast-islands, center-northwest, center-northeast, center-center, center-south, center-islands, south-northwest, south-northeast, south-center, south-islands). In the last two rows we show the implied ratios and the selection on the unobservables that would be needed to drive our results to zero and the value of the coefficient if selection of the observable was identical to the one on the unobservables. The standard errors are shown in parenthesis. The left parenthesis shows robust standard errors, while the right shows two-way clustered standard errors using region of origin and destination as groups.

Table 5: Results of the first stage and IV.

	FIRST STAGE	REDUCED FORM	2SLS
	ΔlogMuseums	logTourist flows	logTourist flows
	(1)	(2)	(3)
Log Noble families_d (pc) - Log Noble families_o (pc)	0.318	0.073	
	(0.010) (0.026)	(0.031) (0.047)	
Log Museums_d (pc) - Log Museums_o (pc)			0.229
			(0.109) (0.086)
Log Population_o	0.557	0.109	-0.018
	(0.016) (0.052)	(0.043) (0.038)	(0.073) (0.054)
Log Population_d	-0.557	0.655	0.782
	(0.016) (0.052)	(0.046) (0.103)	(0.073) (0.101)
Log Distance	0.000	-0.654	-0.654
	(0.011) (0.018)	(0.055) (0.103)	(0.035) (0.083)
Log Regional Income_o (per capita)	-2.889	-0.812	-0.151
	(0.101) (0.279)	(0.313) (0.230)	(0.476) (0.366)
Log Regional Income_d (per capita)	2.889	-1.496	-2.156
	(0.101) (0.279)	(0.292) (0.547)	(0.476) (0.617)
Log Education_o	1.042	2.540	2.302
	(0.330) (0.857)	(1.030) (0.480)	(1.057) (0.331)
Log Education_d	-1.042	-5.803	-5.565
	(0.330) (0.857)	(1.107) (2.247)	(1.057) (2.017)
Log Foreign Tourists_o (per capita)	0.267	0.193	0.132
	(0.023) (0.063)	(0.072) (0.058)	(0.076) (0.055)
Log Foreign Tourists_d (per capita)	-0.267	0.852	0.913
	(0.023) (0.063)	(0.065) (0.103)	(0.076) (0.077)
Observations	380	380	380
R-squared	0.979 0.979	0.869 0.869	0.805 0.805

First stage results using the instrumented variable *Log Museums_d (per capita) - Log Museums_o (per capita)* as dependent variable and the instrument (*Log Noble families_d (per capita) - Log Noble families_o (per capita)*) as an independent variable. Reduced form results using the instrument (*Log Noble families_d (per capita) - Log Noble families_o (per capita)*) as a regressor. We perform a Hausman test, where the null hypothesis is that OLS estimates are identical to the IV ones, and we do not find evidence of endogeneity. Standard errors are in parentheses. The left parenthesis shows robust standard errors, while the right shows two-way clustered standard errors using region of origin and destination as groups.

Table 6: Robustness checks: other specifications

	Log Tourist flows_od (per capita)			
	(1)	(2)	(3)	(4)
Weighted for the population in the region of origin	0.461	1.481	1.478	1.732
	(0.100) (0.116)	(0.260) (0.155)	(0.246) (0.152)	(0.279) (0.176)
Observations	380	380	380	380
R-squared	0.890	0.898	0.903	0.932
Not using per capita values	0.145	0.715	0.714	0.703
	(0.116) (0.222)	(0.284) (0.294)	(0.281) (0.305)	(0.328) (0.441)
Observations	380	380	380	380
R-squared	0.899	0.901	0.903	0.927
Controlling for international flight passengers in the region of origin and destination	0.851	1.710	0.733	2.623
	(0.421) (0.218)	(0.511) (0.312)	(0.685) (0.331)	(0.603) (0.187)
Observations	240	240	240	240
R-squared	0.823	0.848	0.853	0.890
Number of bilateral area dummies	0	3	8	24

OLS estimates. Robustness checks using the number of museums in the region of origin and destination taken separately, weighting for the population in the region of origin, not using per capita values (both in the dependent variable and in the regressors) and, finally, controlling for the number of international flight passengers in the region of origin and destination. The number of observations when we control for international flight passengers is lower than 380 because four regions do not have airports (Basilicata, Molise, Trentino Alto Adige and Valle d' Aosta) and are excluded given the log specification. Column 1 shows results not controlling for bilateral area dummies. Column 2 controls for 3 bilateral area dummies (north, south). Column 3 controls for 8 bilateral area dummies (north, center and south). Column 4 controls for 24 bilateral area dummies (Northeast, Northwest, Center, South, Islands). The standard errors are shown in parenthesis. The left parenthesis shows robust standard errors, while the right shows two-way clustered standard errors using region of origin and destination as groups.

Table 7: Robustness checks: other measures of "culture".

	Log Tourist flows_od (per capita)			
	(1)	(2)	(3)	(4)
Measure of museums taken from "museionline.it"	0.282	0.283	0.336	0.539
	(0.065) (0.082)	(0.072) (0.080)	(0.085) (0.067)	(0.108) (0.086)
R-squared	0.875	0.877	0.881	0.915
A measure of museums' quantity and quality taken from "tripadvisor.com"	0.237	0.289	0.313	0.473
	(0.053) (0.055)	(0.070) (0.069)	(0.069) (0.082)	(0.070) (0.069)
R-squared	0.875	0.879	0.882	0.920
Cultural Index	0.260	0.261	0.263	0.371
	(0.045) (0.049)	(0.047) (0.051)	(0.048) (0.042)	(0.064) (0.064)
R-squared	0.879	0.881	0.884	0.917
Number of bilateral area dummies	0	3	8	24
Observations	380	380	380	380

OLS estimates. Robustness checks using different measures of museums: another measure of the number of museums taken from the website "http://www.museionline.it" and a measure of the quantity and quality of museums (the ranking of cultural attractions in the website "http://www.tripadvisor.it"). Finally we generate a composite index (*the cultural index*), that is an aggregated measure of three different cultural goods (museums, theatrical performances, concerts). Column 1 shows results not controlling for bilateral area dummies. Column 2 controls for 3 bilateral area dummies (north, south). Column 3 controls for 8 bilateral area dummies (north, center and south). Column 4 controls for 24 bilateral area dummies (Northeast, Northwest, Center, South, Islands). Standard errors are shown in parenthesis. The left parenthesis shows robust standard errors, while the right shows two-way clustered standard errors using region of origin and destination as groups.

A Data and descriptive statistics

Data on *Tourist flows* measure the number of Italian tourists who paid for an accommodation at least one night in a region which is not their own region. Data are taken from "Arrivi e presenze degli italiani negli esercizi ricettivi per regione di provenienza e di destinazione" of the Italian Statistics Bureau (ISTAT, 2006). *Foreign tourists* (in thousands) are the the number of foreign tourists who spent at least one night in an accommodation facility in an Italian region (source: Osservatorio nazionale del Turismo, 2006). Data on *Museums* are extracted from "I musei e gli istituti similari non statali" (ISTAT, 2006), from "Visitatori e introiti di Musei, Monumenti e Aree Archeologiche Statali - Dati per Provincia e Regione" (Ministero dei Beni e delle Attività Culturali e del Turismo - Ufficio Statistica, 2006" and from the websites "Museionline" (http://www.museionline.it) and "TripAdvisor" (http://www.tripadvisor.it). Data on *Population* (expressed in 100,000 inhabitants) come from "Indicatori demografici", (ISTAT, 2006). *Regional income* is expressed in billions of Euros and is taken from "Conti economici regionali-Valore aggiunto ai prezzi base e prodotto interno lordo" (ISTAT, 2006). The regional *Gini Index* is computed on disposable net household income and is taken from "Diseguaglianza dei redditi per regione-Indice di concentrazione di Gini sui redditi netti familiari esclusi i fitti imputati" (ISTAT, 2007). The *Dependency ratio* is the ratio between the population aged 65 or over and the population aged 20-64, while *Education* is measured as the percentage of population with at least a middle school diploma. The source of these statistics is "Health for all – Italia" (ISTAT, 2006). The *Consumer Price Index* is a proxy for the cost of living at regional level (ISTAT). *Geographic distance* between regions is measured as the distance between the region capital cities (in hundreds of km). To collect these data we used the Italian Road Atlas (Touring Club Italiano, 2004). *Parks* is the ratio between the surface covered by parks and the total surface in each region. Data come from the "Direction for the Nature Protection" (Ministry of Environment and Natural Resources Protection, 2003). *Coasts* are measured as the ratio between the coastline length of each region and the total coastal length of Italy. The source of the data is "Istituto Superiore per la Protezione e la Ricerca Ambientale". *Mountains* are measured as the ratio between mountain areas and total surface in each region. The source is the database "Unione Nazionale Comuni, Comunità, Enti Montani" of the Ministry of Agriculture and Forestry. Data on the regions with ski resorts come from the website www.http://regioni-italiane.com. Data on *Concerts* and on *Theatrical performances* come from "I dati dello Spettacolo" (Società Italiana degli Autori e degli Editori, 2006), while Data on the presence of Italian noble families are taken from "Libro d'Oro della Nobiltà Italiana" (Collegio Araldico, 2004) (in English: "Golden Book of the Italian Nobility"), that is regularly published by the Collegio Araldico of Rome. It lists most of Italy's noble families. It was first published in 1910 and it includes those families listed in the register of the "Libro d'Oro della Consulta Araldica del Regno d'Italia and in the "Elenchi Ufficiali Nobiliari" (both of the year 1921 and 1933). The book is a comprehensive listing of families that are considered noble in Italy.

B OLS estimates with all the regressors

Table 8: Estimates of the OLS regressions with all the regressors.

	Log Tourist flows_od (per capita)			
	(1)	(2)	(3)	(4)
ΔlogMuseums	0.383	1.469	1.473	1.829
	(0.086) (0.108)	(0.219) (0.088)	(0.219) (0.094)	(0.259) (0.195)
Log Population_o	-0.099	-0.802	-0.790	-0.772
	(0.057) (0.080)	(0.138) (0.083)	(0.138) (0.105)	(0.139) (0.117)
Log Population_d	0.863	1.539	1.568	1.834
	(0.060) (0.124)	(0.141) (0.068)	(0.140) (0.069)	(0.142) (0.130)
Log Distance	-0.654	-0.723	-0.701	-0.734
	(0.054) (0.092)	(0.064) (0.096)	(0.079) (0.115)	(0.057) (0.097)
Mountain_o	-0.598	0.522	0.538	1.662
	(0.200) (0.144)	(0.276) (0.236)	(0.277) (0.318)	(0.444) (0.219)
Mountain_d	-1.540	-2.412	-2.385	-2.468
	(0.214) (0.366)	(0.288) (0.368)	(0.286) (0.352)	(0.473) (0.379)
Ski_o	-0.564	0.063	0.122	-0.891
	(0.170) (0.086)	(0.197) (0.084)	(0.208) (0.108)	(0.281) (0.191)
Ski_d	-0.605	-1.181	-1.066	-1.480
	(0.169) (0.217)	(0.206) (0.122)	(0.213) (0.130)	(0.296) (0.266)
Mountain x Ski_o	1.393	0.060	0.045	2.376
	(0.330) (0.165)	(0.407) (0.205)	(0.410) (0.286)	(0.637) (0.422)
Mountain x Ski_d	0.814	1.970	1.953	2.836
	(0.329) (0.478)	(0.420) (0.262)	(0.417) (0.256)	(0.645) (0.580)
Park_o	0.200	3.895	3.653	3.649
	(0.520) (0.638)	(0.804) (0.501)	(0.809) (0.541)	(0.805) (0.474)
Park_d	1.315	-2.266	-2.786	-4.185
	(0.568) (0.889)	(0.888) (0.307)	(0.886) (0.442)	(0.835) (0.612)
Coast_o	-1.952	6.801	6.527	10.057
	(0.917) (0.918)	(1.689) (0.755)	(1.685) (0.977)	(2.411) (0.838)
Coast_d	-2.684	-10.727	-11.305	-13.611
	(0.899) (1.413)	(1.700) (1.001)	(1.695) (1.099)	(2.341) (1.623)
Log Regional Income_o (per capita)	0.352	2.491	2.494	2.674
	(0.399) (0.454)	(0.572) (0.243)	(0.566) (0.237)	(0.576) (0.224)
Log Regional Income_d (per capita)	-2.659	-5.131	-5.166	-6.123
	(0.410) (0.716)	(0.571) (0.434)	(0.549) (0.232)	(0.548) (0.350)
Log Education_o	2.272	-1.422	-1.751	3.522
	(1.004) (0.580)	(1.242) (0.548)	(1.335) (0.674)	(1.582) (0.550)
Log Education_d	-5.535	-2.639	-3.311	-1.225
	(1.094) (2.107)	(1.260) (0.851)	(1.297) (1.009)	(1.565) (1.476)
Log Foreign Tourists_o (per capita)	0.094	-0.545	-0.520	-1.141
	(0.073) (0.091)	(0.124) (0.078)	(0.124) (0.096)	(0.230) (0.116)
Log Foreign Tourists_d (per capita)	0.950	1.525	1.580	1.672
	(0.069) (0.110)	(0.126) (0.077)	(0.125) (0.065)	(0.221) (0.167)
CPI_o	-0.001	-0.059	-0.048	-0.102
	(0.009) (0.013)	(0.013) (0.008)	(0.017) (0.017)	(0.023) (0.015)
CPI_d	0.104	0.162	0.185	0.197
	(0.009) (0.015)	(0.013) (0.014)	(0.015) (0.017)	(0.022) (0.024)
Gini Index_o	-1.033	-5.652	-4.543	-19.748
	(2.477) (2.052)	(2.483) (1.290)	(2.681) (1.414)	(4.747) (2.230)
Gini Index_d	11.770	17.079	19.451	18.609
	(2.723) (5.133)	(2.550) (2.426)	(2.729) (2.902)	(4.754) (4.729)
Dependency Ratio_o	0.033	0.023	0.024	0.038
	(0.012) (0.014)	(0.014) (0.019)	(0.014) (0.020)	(0.012) (0.011)
Dependency Ratio_d	0.003	-0.002	0.001	-0.000
	(0.012) (0.017)	(0.014) (0.017)	(0.014) (0.015)	(0.012) (0.012)
Number of bilateral area dummies	0	3	8	24
Observations	380	380	380	380
R-squared	0.874	0.886	0.889	0.921

Regression results with all the regressors that we use in our specification. Column 1 shows results not controlling for bilateral area dummies. Column 2 controls for 3 bilateral area dummies (north, south). Column 3 controls for 8 bilateral area dummies (north, center and south). Column 4 controls for 24 bilateral area dummies (Northeast, Northwest, Center, South, Islands). The standard errors are shown in parenthesis. The left parenthesis shows robust standard errors, while the right shows two-way clustered standard errors using region of origin and destination as groups.