Migrant Networks and the Spread of the Coronavirus in Italy[†]

Vincenzo Caponi, University of Sassari, CREST,IZA

April 4, 2020

Abstract

The focus of this ongoing project si to look at migrant networks and the possible role that might have in the spreading of the virus covid-19 in Italy. I propose a spatial econometric model with a distance matrix that proxies the "social" proximity of Italian provinces, to look at the "spillover effects" that the government policies have caused in terms spreading of the virus. The purpose of the paper is two-fold: on the one hand it aims at showing the possible endogeneity of the policy instruments used highlighting the counter-productive effects obtained by neglecting the response of individuals; on the other hand it aims at measuring those effects by considering a novel and so far unused measure of network that is not observed producing relevant effects in "normal" times, but it does in special circumstances. Networks are used widely to implement simulations of virus spreading, but they are mostly approximated by data on human movements in "normal" times, such as number of flights by route, number of commuters taking the trains etc..., and are most often assumed to be exogenous to policy implementation. In this paper I show that other types of networks that generate less physical interactions in normal times, such as migrant networks, in special circumstances can find themselves "activated" and produce important effects.

JEL Classification: F22, J61, O15

[†]This is an ongoing project. Given that this is no time for seminars and presentations, I chose to make it public in case is of interest to anyone and have feedback and questions. In case of citation it would be better to contact me first to agree on points that are very unlikely to be substantially revised.

1 Introduction

Italy has been the first European country to experience the large-scale spread of the coronavirus. The spread at first was concentrated in some Northerner small districts, reaching gradually larger cities such as Milan. The Government managing reaction was at first to try to contain the contagion by isolating those smaller areas, and implementing locally severe measures of social distancing. On March 7 the Government announced that those measures were going to be extended to the whole region of Lombardy and other 14 large provinces of the North, starting on March 8. The announcement was also preceded by rumors that this was going to happen. Right after the rumors spread and even more after the Government announcement, given that the policy was still not in effect, a large traffic of people was recorded going from the North to the South of Italy, by trains, planes and cars. Some also by ferries to Sicily and Sardinia. On March 9, the Government decided to extend the "red-zone" to the whole country starting from March 10, another intensification of the traffic North-South was also recorded during the night between 9 and 10 of march. As of April 3rd, looking at data by province of contagion and death rates in Italy, we can clearly see an increase of cases in the Southerner provinces that exceeds in most cases that of the North. While the absolute number remains lower, there is an increased preoccupation that this number can climb fast, creating a pressure to the health system in the south that could be harder to manage compared to the North. The focus of this study is to look at migrant networks, as defined by clusters of citizens of the Italian peninsula that from one province move to another, and the possible role that have in the spreading of the virus, given the announces made in the days between the 7 and the 10 of march. I draw from the literature in regional and international economics and spatial econometrics proposing a model of spillover "effects" caused by the reaction to the government policy of announcing the lock-down of large regions of Italy before implementing strict measures to prevent movements. The main idea of the paper is that the flow of people from the North to the South could have been anticipated looking at social vicinity between small areas on the two sides of Italy, where for social vicinity I intend the presence of clusters of immigrants from southerner areas in northern regions. I create a distance matrix that proxies the social vicinity in terms of networks, and use this matrix to weigh observations of cases of contagion per millions inhabitants in the Northern regions, and check the correlation of this variable with cases in Southern regions 5-6 days after. Indeed, I find a strong and positive correlation. I also try two other measures of proximity based on geography, the distance between the capital oc the provinces in kilometers and the time it would take to make a trip by car, both variables calculated using google maps, and I find that they do not have an important role in explaining the diffusion of the virus from North to South, indeed, they show a negative correlation. This paper contributes to the epidemiological literature, especially on spreading of SARS viruses, taking one step further the concept of networks in metapopulation models.[1]

2 Data and Facts

As of April 3rd the covid-19 virus has infected more than a million people around the world and caused more than 60,000 deaths. Most of the infected were frist in China, with more than 80,000 cases. However China seems to have contained the virus in the province of Hubei, and in particular in the city of Wuhan and it sees its cases not growing any further. Italy was the second country to see the virus fast spreading and differently from Chian seem to have failed to contain the virus where first appeared. As of April 3rd the infections recorded in Italy are more than 115,000 with almost 15,000 deaths. Here I use data made available daily by the "Protezione Civile" publicly on gitub. Additionally, for population and migration rates I use data from Istat by province. Figure 1 shows the development of infected cases in the Italian provinces. The maps in the figures represent the numbers at six different dates for the south while in the North the cases are kept at the 17th of Februry. What can be seen clearly is that while originally the infection was only in certain areas of the North, by April 3rd most of the South is also concerned. However, not all provinces in the South were affected with to the same extent. Some provinces of Puglia for example were more affected than others and in general more than other Southern provinces except some of Sicily and the province of Sassari. It is quite reasonable to assume that the virus moved from North to South with people, but data on this movement are not available at the moment. Anecdotal evidence suggests that many people, once schools and many places of work closed down in the Norther regions, decided to move to the South, reports by news papers indicate that when the closing of major area was announced thousands of people crowded train stations and airports directed to Southern cities. Governors of Southern regions also alarmed the central government that they were seeing an abnormal flow of people coming from North. But who moved where? In absence of data on the actual flows we can speculate that many of those people had some relationship with the areas of the South where they moved to, and this relationship could well be family based. In recent years many Southerners moved to North, this migration flow has been substantial, hence the North is populated by large clusters of Southern immigrants. It is therefore natural to think that those people are the ones that most likely moved "back" to the south. Going back to Figure 1 March 17 for the North was chosen as the median number of days for symptoms to appear and therefore for cases to be registered is about 5 to 7 days,

[1] See Colizza and Vespignani (2008) for more details.

hence March 17 seems to be a good date to approximate what the situation in terms of potential risk of contagion in the North was at the time many people left for the South. I chose to report only the first 15 provinces in terms of number of cases, the most affected. In the south are reported all provinces with a positive number of cases, those provinces are divided in to two groups in terms of social proximity. The orange bubble represent the provinces that are closer to at least one of those 15 provinces in the North, that is, there is a large enough community originating in the southern province in one of the north, the blue bubbles are provinces that are not as close. As we can see up to March 10 there are virtually no cases in the South, on March 17 we see some provinces with a very small number of cases, by April 2nd the virus is clearly present all over the South.

It is interesting to notice that in the area of Campania all bubbles are blue, while in Sicily, Sardinia and Apulia there are more orange bubbles. The 24th of March sees a significant increase of cases especially in the orange provinces which is even more accentuated on April 2nd. Orange bubble grow faster, but are also geographically more distant from the northern provinces with more cases. It is therefore clear that geographical distance cannot explain this pattern by itself. Moreover, it is also quite interesting to notice that the dates around March 7 to 10 are important to explain the diffusion mechanism. The following section will analyse these hypotheses more in details, with a spatial econometric model.

3 Model and Results

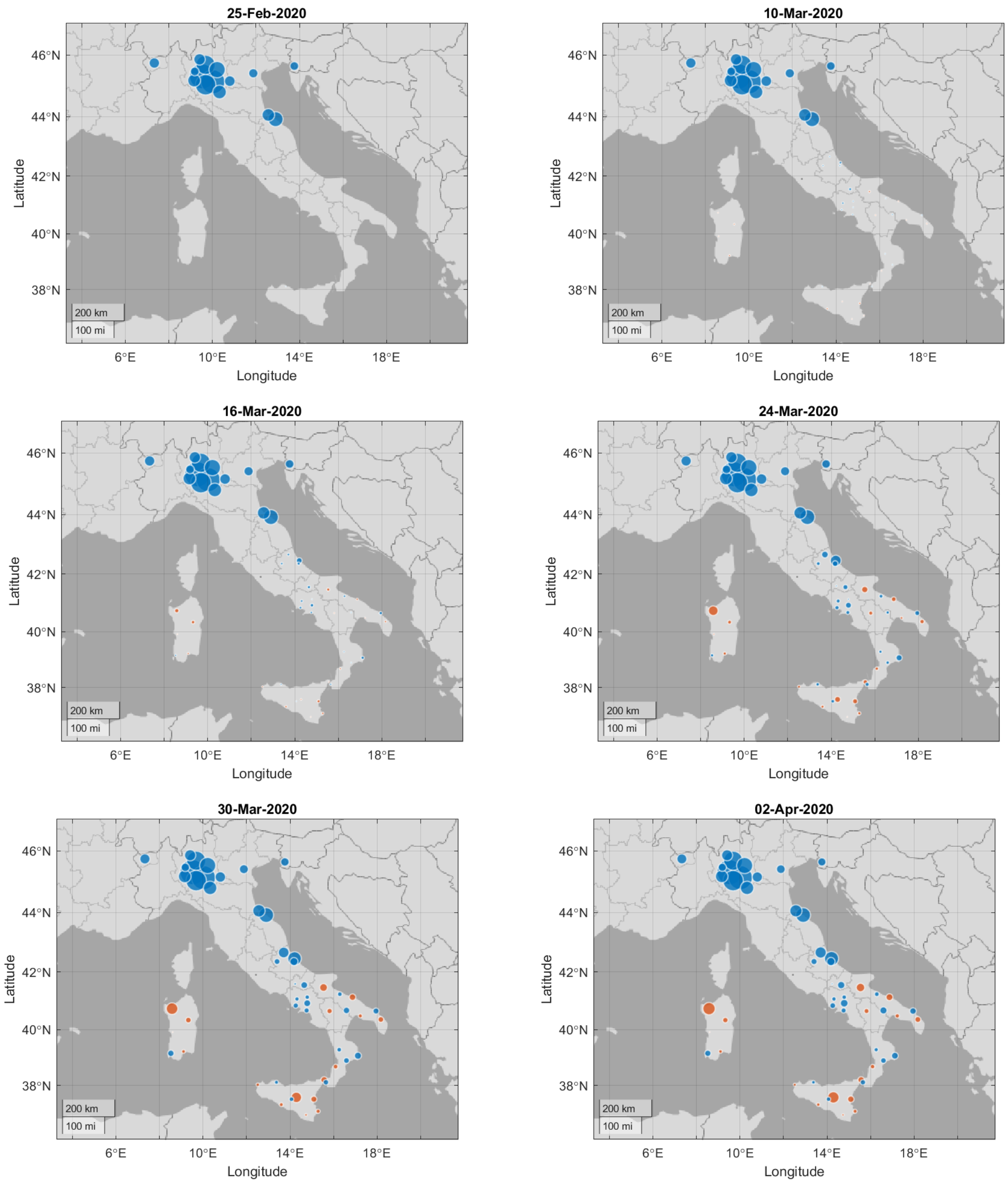
The model I used is a modification of the workhorse model in spatial econometrics for estimating "spillover" effects. The data available have a panel structure, as I have 107 provinces observed for daily from February 24 to April 2. I stack all the observations to obtain a vectored representation of the panel, and I divide the provinces in terms of South and North by splitting the sample in macro-regions. I take the log of the number of total infected cases per 100,000 people as the series to study. The model is then,

Having said this, the model estimated is

$$Y_t = \alpha_j Y_{t-j} + \beta X + \theta_1 W_t Y_t + \theta_2 W_t X + \theta_{3j} W_t Y_{t-j} + \epsilon_t \tag{1}$$

where W is the spatial matrix of weights that may be time dependent. Essentially the model boils down in an space-time autoregressive model, or a SpARMA model in which the moving average component is set equal to zero. The matrix X captures fixed effects that are explained by a series of time-invariant variables, such as population density in a province, the ratio of the population that is

Figure 1: The Spreading of the Virus from North to South



older that 85 years, and the share of families with kids, all variables that are assumed to be positively correlated with the spread of the virus. I compare the results obtained by using alternative weighting matrices, in particular I use classical matrices based on geographical distances, measured in kilometers or time to cover the distance by car (both taken from google), and two measures of distances based on immigrants flows. I also take those matrices setting all zeros for within macro-regions distances in order to look more specifically to the North-South spillover effect.[2]

3.1 Results

Table **??** shows the results of the first SDM model. The first panel shows the coefficient estimates and their te-test values, while the second panel the direct, indirect and total effects.[3] The Model is run with three weighing matrix, the first is the migrant network matrix and the other two the actual distances in terms of kilometers and car-hours. The estimates of the three models are very similar in that they reproduce a spatial autoregressive parameter of similar and important magnitude and highly significant. The coefficient of the lagged dependent variable is also very significant and close to 1, which suggests a high autocorrelation in time and that associated to the weighted lagged dependent variable is negative in all models. Population density seems not to have a role in explaining the different levels of contagion across provinces. To understand better the role of the weighing matrix we can look at Table **??** in which we can read the direct, indirect and total effects. Here we see that all three models show important indirect - "spillover" - effects, however model 1 shows indirect effects much larger than the other two models, given that the weighing matrices are all normalized, this suggests that social proximity as measured by immigrant networks is more important for the transmission of the contagion that geographical proximity.

Table 3.1 and Table 3.1 report the estimates of the models 1 to 3 for the log-differences of cases per population. It is interesting here to see that for model 1 the coefficient on the lagged dependent variable is not significant while it is highly significant the coefficient of that variable weighted with the social proximity matrix. The other two models have more ambiguous results. The spatial autoregressive coefficient is significant in all models and smaller in the first model, however, looking at the indirect effects we have that the ones associated with the lagged dependent variable are higher for the first model than the others, although probably not significantly, and are also much more precisely estimated for model 1. Indeed, the total effect of the lagged variable is suggested to be coming mainly from spillover effects rather than directed ones.

[2]It turns out that the W letting W be time dependent does not improve the predicting power of the model, hence I work with a constant W in what follows.

[3]See LESAGE (2008) for all the details about the model and also the Matlab code that I use.

Spatial Durbin Model - differences - w/o Policy

Dependent Variable: Cases x 100,000 population

Parameter	Coefficient	Std. Error	t-stat
Model			
Parameter	Coefficient	Std. Error	t-stat
Lagged Y	0.05112	0.04463	1.14563
Pop. Density	0.00205	0.01594	0.12843
W*Lagged Y	0.40455	0.01099	36.80426
W*Pop. Density	0.00190	0.04624	0.04100
ρ	0.38798	0.01057	36.70130
R^2	0.02038		
Model			
Parameter	Coefficient	Std. Error	t-stat
Lagged Y	0.06445	0.19665	0.32775
Pop. Density	0.00075	0.00407	0.18348
W*Lagged Y	0.21297	0.05187	4.10584
W*Pop. Density	0.00746	0.06767	0.11030
ρ	0.56997	0.14038	4.06006
R^2	0.02711		
Model			
Parameter	Coefficient	Std. Error	t-stat
Lagged Y	0.06088	0.00200	30.47457
Pop. Density	0.00076	0.01914	0.03963
W*Lagged Y	0.24046	0.04109	5.85219
W*Pop. Density	0.00681	0.10954	0.06216
ρ	0.55597	0.09280	5.99108
R^2	0.02790		
N.Obs:	CS: 107	TS 30	

Average Marginal Effects - differences - w/o Policy

Dependent Variable: Cases x 100,000 population

Parameter	Coefficient	Std. Error	t-stat
Marginal Average Effects Model 1			
Direct Effects			
Lagged Y	0.06283	0.01604	3.91637
Pop. Density	0.00208	0.00087	2.39439
Indirect Effects			
Lagged Y	0.68519	0.06450	10.62379
Pop. Density	0.00418	0.00238	1.75343
Total Effects			
Lagged Y	0.74802	0.06351	11.77858
Pop. Density	0.00625	0.00259	2.41149
Marginal Average Effects Model 2			
Direct Effects			
Lagged Y	0.06894	0.01519	4.53875
Pop. Density	0.00087	0.00090	0.97427
Indirect Effects			
Lagged Y	0.57593	0.14665	3.92719
Pop. Density	0.01835	0.00776	2.36290
Total Effects			
Lagged Y	0.64487	0.14762	4.36843
Pop. Density	0.01922	0.00779	2.46695
Marginal Average Effects Model 3			
Direct Effects			
Lagged Y	0.06676	0.00424	15.75955
Pop. Density	-0.00023	0.02134	-0.01092
Indirect Effects			
Lagged Y	0.67594	0.33684	2.00669
Pop. Density	-0.04968	0.34565	-0.14372
Total Effects			
Lagged Y	0.74270	0.34060	2.18057
Pop. Density	-0.04991	0.35916	-0.13896

These results are already quite interesting as they clearly say that social proximity has a role in explaining the contagion process as much as geographical proximity. However, geography and migration are certainly not uncorrelated and social proximity could indeed pick up the effect that geography has. Trying to disentangle the two effects I run the models recalculating the weighing matrices setting equal to zero all the cells that correspond to provinces in the same macro areas, where macro areas are North and South of Italy.[4] Tables **??** and **??** show the results for the models in differences. What we can see here is that the weighted lagged dependent variable is positive and significant in model 1 and negative and significant in models 2 and 3. The spatial autocorrelation coefficient is still lower in model 1, but turning to the average marginal effects we can see that the spillover effects are large, positive and significant in model 1, while they are negative in models 2 and 3 and significant only in model 3. In other words, when we take out the more obvious effect of geographical distance between neighbouring provinces and provinces that are at the other end of Italy, the spillover effects vanish, instead the social proximity still plays a significant role.

Finally, I introduce a dummy variable for policy that is equal to 1 for the period following March 10, when the government has announced the lock down of the northern regions first and, the day after, of the whole country. The idea is to see if there is a before and an after these events. Table 3.1 reports the results. I use again the corrected matrix that has all zeros for same macro area provinces.

Focusing on Model 1, the dummy policy has a negative and significant coefficient when not weighted and becomes positive and also significant when weighted. This clearly suggests that the lock down has a mitigating direct effect, but it did accelerate the spreading of the virus, especially from the North to the South. Looking at the average marginal effects clearly shows that the dummy policy increased the spillover effects along the lines of the social proximity.

4 Simulating the Contagion

In this section I first complement the model with a third lag of the level of the endogenous variable and a second autoregressive lag of the dependent variable in order to improve the predicting power of the model. I also try different timing of the first policy dummy as it is difficult to state clearly when the policy may show its effects, and I choose the timing that gives the best prediction in terms of the lowest sum of squares of the distance between the national number of fitted and actual total cases. The results of the model estimates are shown in Table **??**. The model uses the weights from the social proximity matrix are used and only those representing the South to North flow.

The Figures 2 to 3 show the predicting performance of the model for the whole country and for

[4] I take the usual division that says Lazio, Umbria and Marche are the bordering northern regions with the South.

Spatial Durbin Model - diff - ns - w/o Policy

Dependent Variable: Cases x 100,000 population			
Model			
Parameter	Coefficient	Std. Error	t-stat
Lagged Y	0.16315	0.05246	3.11013
Pop. Density	0.00508	0.01639	0.30993
W*Lagged Y	0.24604	0.00844	29.16451
W*Pop. Density	0.00767	0.05215	0.14708
ρ	0.20997	0.00732	28.67620
R^2	-0.03630		
Model			
Parameter	Coefficient	Std. Error	t-stat
Lagged Y	0.17322	0.15053	1.15073
Pop. Density	0.00488	0.01597	0.30594
W*Lagged Y	-0.13976	0.00374	-37.38278
W*Pop. Density	0.06695	0.13552	0.49404
ρ	0.35100	0.01086	32.33102
R^2	-0.06007		
Model			
Parameter	Coefficient	Std. Error	t-stat
Lagged Y	0.17971	0.00764	23.53024
Pop. Density	0.00541	0.01600	0.33801
W*Lagged Y	-0.09442	0.00267	-35.35972
W*Pop. Density	0.06422	0.11933	0.53812
ρ	0.36899	0.01040	35.49131
R^2	-0.06804		
N.Obs:	CS: 107	TS 30	

Average Marginal Effects - diff - ns - w/o Policy

Dependent Variable: Cases x 100,000 population

Parameter	Coefficient	Std. Error	t-stat
Marginal Average Effects Model 1			
Direct Effects			
Lagged Y	0.16704	0.05204	3.20984
Pop. Density	0.00563	0.01611	0.34926
Indirect Effects			
Lagged Y	0.35355	0.01859	19.01521
Pop. Density	0.01141	0.06879	0.16585
Total Effects			
Lagged Y	0.52059	0.06458	8.06139
Pop. Density	0.01703	0.07165	0.23775
Marginal Average Effects Model 2			
Direct Effects			
Lagged Y	0.17445	0.14793	1.17923
Pop. Density	0.00466	0.01622	0.28750
Indirect Effects			
Lagged Y	-0.12246	0.08012	-1.52847
Pop. Density	0.11127	0.21121	0.52683
Total Effects			
Lagged Y	0.05199	0.22788	0.22813
Pop. Density	0.11594	0.21312	0.54400
Marginal Average Effects Model 3			
Direct Effects			
Lagged Y	0.17958	0.00772	23.26271
Pop. Density	0.00452	0.01652	0.27385
Indirect Effects			
Lagged Y	-0.04459	0.00542	-8.23230
Pop. Density	0.09897	0.18838	0.52537
Total Effects			
Lagged Y	0.13499	0.01261	10.70229
Pop. Density	0.10349	0.19022	0.54406

Spatial Durbin Model - diff - Wns - with Policy

Dependent Variable: Cases x 100,000 population			
Model 1			
Parameter	Coefficient	Std. Error	t-stat
Lagged Y	0.19629	0.00238	82.64682
dummy policy	-0.13766	0.02039	-6.75191
Pop. Density	0.00483	0.03509	0.13779
W*Lagged Y	0.16336	0.00488	33.45865
W*dummy policy	0.17386	0.05843	2.97568
W*Pop. Density	0.00856	0.05838	0.14667
ρ	0.21799	0.00658	33.15460
Marginal Average Effects Model 1			
Direct Effects			
Lagged Y	0.19742	0.00240	82.14764
dummy policy	-0.13561	0.02007	-6.75734
Pop. Density	0.00300	0.03519	0.08515
Indirect Effects			
Lagged Y	0.26224	0.00986	26.59471
dummy policy	0.18142	0.07174	2.52867
Pop. Density	0.01618	0.06925	0.23371
Total Effects			
Lagged Y	0.45966	0.01033	44.47664
dummy policy	0.04581	0.07453	0.61463
Pop. Density	0.01918	0.06484	0.29579

the South and North of Italy separated. As we can see the prediction seems to be very accurate for the whole country and for the North, for the South the model performs with more uncertainty for the starting period. This is likely the results that the South presents many zeros at the beginning of the time period, hence the mode is probably not the best to fit these data, reinvitation of some sort of a tobit may be more appropriate. However, for later periods the model perform well, and this is undoubtedly the strength of the geographical component of the model. That is, even if the South starts with many zeros, it clearly picks up contagion due to the social proximity with the North.

Overall, not only the model can predict the country wide number of cases, but also the regional pattern very well. We can also clearly notice from the graphs that the policy variable induces much of the concavity of the predicting line. That means that the effects that we started to see on March 15 of the lock down policy are important, although they are less evident for the South as in this region the policy has a counter-acting spillover effect.

5 Conclusions

This is a very preliminary draft and any conclusion must be taken with a grain of salt. However, the main message of this work is that migration networks may have plaid an important role in the diffusion of the covid-19 virus across Italy interacting with the policy set up by the government which locked down the whole country, but it left a few hours or days for people to move from region to another between the announcement of the policy and the time it was effective. Similar situations

Spatial Durbin Model - differences - w/o Policy

Dependent Variable: Cases x 100,000 population			
Model			
Parameter	Coefficient	Std. Error	t-stat
Constant	0.1441	0.1074	1.3421
Lag 1 Δ Y	0.0178	0.0085	2.0943
Lag 2 Δ Y	0.0737	0.0159	4.6346
Lag 3 Y	0.0106	0.0162	0.6566
Pol. Lag 1 Δ Y	0.0444	0.0036	12.4216
Pol. Lag 2 Δ Y	-0.0081	0.0325	-0.2495
Pol. Lag 3 Y	-0.0254	0.0312	-0.8132
W*Lag 1 Δ Y	-0.4343	0.0185	-23.5046
W*Lag 2 Δ Y	-0.3211	0.0897	-3.5814
W*Lag 3 Y	0.0772	0.0982	0.7860
W*Pol. Lag 1 Δ Y	1.0351	0.0463	22.3674
W*Pol. Lag 2 Δ Y	0.2004	0.3075	0.6516
W*Pol. Lag 3 Y	-0.0864	0.2877	-0.3003
ρ	-0.1450	0.0113	-12.8277
R^2	0.08		

Marginal Average Effects Model 1			
Direct Effects			
Lag 1 Δ Y	0.0180	0.0086	2.0961
Lag 2 Δ Y	0.0737	0.0161	4.5902
Lag 3 Y	0.0105	0.0160	0.6531
Pol. Lag 1 Δ Y	0.0444	0.0035	12.5313
Pol. Lag 2 Δ Y	-0.0079	0.0339	-0.2329
Pol. Lag 3 Y	-0.0246	0.0314	-0.7811
Indirect Effects			
Lag 1 Δ Y	-0.3816	0.0143	-26.6550
Lag 2 Δ Y	-0.2872	0.0791	-3.6309
Lag 3 Y	0.0632	0.0840	0.7528
Pol. Lag 1 Δ Y	0.8985	0.0356	25.2457
Pol. Lag 2 Δ Y	0.1562	0.2735	0.5713
Pol. Lag 3 Y	-0.0497	0.2542	-0.1955
Total Effects			
Lag 1 Δ Y	-0.3636	0.0207	-17.5563
Lag 2 Δ Y	-0.2135	0.0821	-2.6004
Lag 3 Y	0.0737	0.0853	0.8637
Pol. Lag 1 Δ Y	0.9429	0.0368	25.5902
Pol. Lag 2 Δ Y	0.1483	0.2762	0.5371
Pol. Lag 3 Y	-0.0743	0.2562	-0.2898

Figure 2: Spatial Autocorrelations in Time

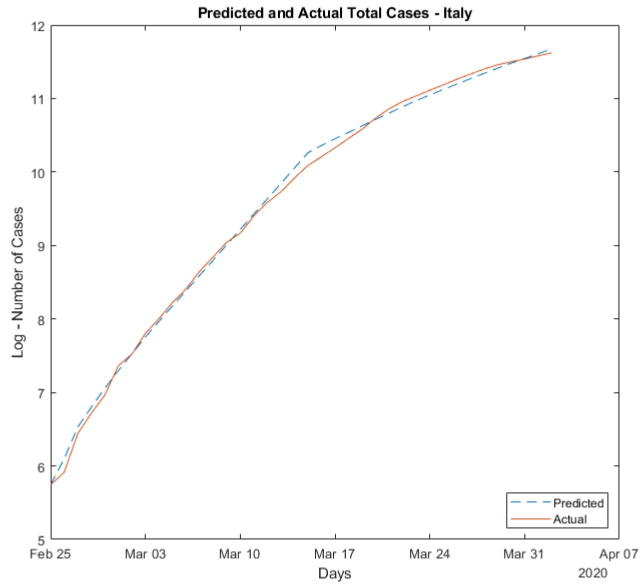
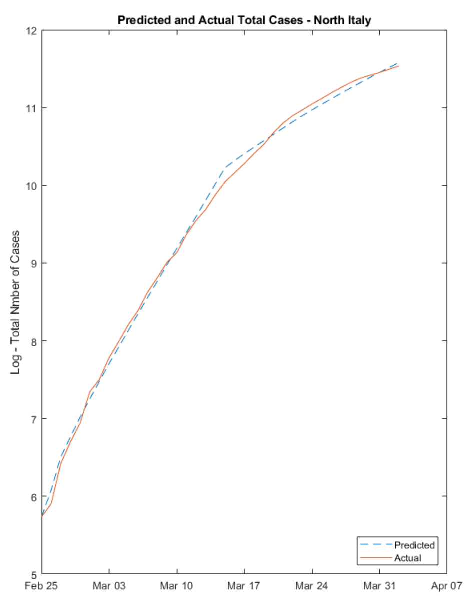
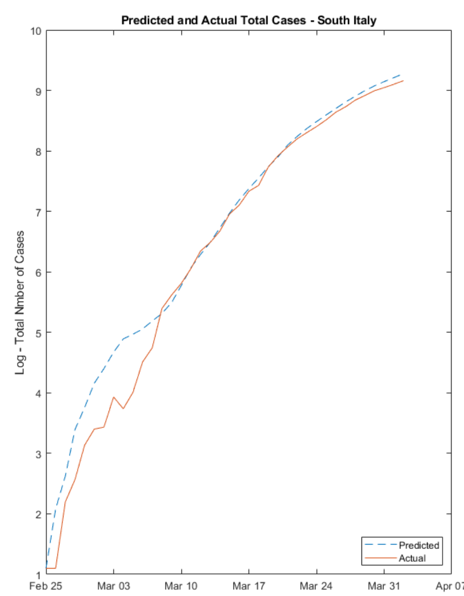


Figure 3: Spatial Autocorrelations in Time





have been registrer in France and in the US, it would therefore be interesting to see if any effect can be seen there too. The importance of this finding is two-fold, it could help in better predicting the effect of such policies for other countries, but it could also give a good indication of how to model a "return to normal" policy. In fact, many of those people who moved south right before or soon after the lock down policy are still there, once the Northern regions will open their factories and schools many may need or want to go back to the North. If meanwhile the contagion has been spread in those communities where they first moved, this can represent a treat to those regions that fought the most to eradicate the virus.

References

Colizza, Vittoria and Alessandro Vespignani, "Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations," *Journal of Theoretical Biology*, 2008, *251* (3), 450 – 467.

LESAGE, James P, "AN INTRODUCTION TO SPATIAL ECONOMETRICS," *Revue d'économie industrielle*, 2008, *123* (3e).