Appendix A Data Description and sources

In this appendix we describe in more detail the variables and their sources. Table 6 provides summary statistics. The first subset of explanatory variables relates to *demographic* characteristics. They include population density in each province, average age, the average size of families, the share of students, the share of secondary school acquisition among 19+ years old residents, the share of postgraduate degree acquisition, the share of families with only one component, and the share of families with five or more components. The Italian national statistical agency ISTAT provides these measures either in 2019 or in 2011, the last year of the full Census. We also create a variable that weighs the number of students with the percent of remote-teaching conducted in each province on 15th November 2020 [8]. The second subset of explanatory variables relates to *economic* characteristics. They include average income per capita (source: Eurostat, 2017), the share of employed workers in the population, share of the agricultural sector, the share of the industrial sector, the share of the service sector, and share of retail and accommodation activities (source: all ISTAT 2019). We also create a variable that weighs the share of retail and accommodation with the percent of businesses that remained open in each province (Ministry of Health) during fall 2020. The third subset of explanatory variables relates to *commuting* activities. We build two measures based on on the total commuting by public transport with trips longer than 15 minutes for i) work, and ii) study reasons. Using the detail of the hour at which commuters leave home and by what transportation mean, we build a measure of iii) concentration of long (¿15 minutes) trips on public transport, weighted by the covid concentration in the province of destination. Finally, we build four measures of exposure through outgoing (OUT) or incoming (IN) commuters to covid. The variables are calculates as

$$X_{ij} = \frac{\sum\limits_{ab \neq ij} C_{ab} \text{flow}^{D}_{(ij)(ab)}}{\sum\limits_{ab \neq ij} \text{flow}^{D}_{(ij)(ab)}} \tag{6}$$

Where ab is any other province different from ij, C_{ab} is the covid incidence per capita in province ab and $\text{flow}^{D}_{(ij)(ab)}$ is the flow from either ij to ab if $D = \text{OUT}$ or from ab to ij if $D = \text{IN}$. In practice, these variables are the average of neighbours' covid incidence, weighted by the commuting flows. These aim to capture whether commuting is a relevant predictor of local covid incidence as a function of whether local commuters work in provinces with high incidence (OUT) or local workers come from provinces with high incidence (IN). We build four variables of this kind: iv) commuting covid IN, v) commuting covid OUT, vi) commuting covid IN (using public transport flows only), and vii) commuting covid OUT (using public transport flows only). The original commuting data are from ISTAT, 2011 Census; we use the official cases in the whole second wave (1/09/2020-23/12/2020) to construct covid exposure. The fourth subset of variables relates to the *health and public health system.* They include mortality rate for cancer in the period 2012-2016, the mortality rate for heart attack in the period 2012-2016, increased life expectancy in the period 2002-2017, asthma incidence, measured as pro-capita consumption of medicine for asthma and Chronic Obstructive Pulmonary Disease (COPD), diabetes incidence, measured as pro-capita consumption of medicine for diabetes, hypertension incidence, measured as pro-capita consumption of medicine for hypertension, the average number of general practitioner doctor per capita, average number of hospital beds per capita. These data are retrieved from the *Health index survey* from *il Sole 24 ore.* The fifth subset of variables includes a *geographical characteristic*: the temperature registered in the period 2007-2016. (source: ISTAT). Finally, we include a measure of covid-19 incidence pre-September 2020, which captures the first wave's strength across provinces.

Hence, the dataset of explanatory variables is composed of 35 variable.

In addition to these, we collect data on the covid-19 incidence between 1/09/2020-3/11/2020, 4/11/2020-23/12/2020, 25/11/2020-23/12/2020, 1/09/2020-26/01/2021, and 26/02/2020-26/01/2020. We do not include these variables in the LASSO selection procedure, as we use them as dependent variables.

Table 4: **Data**

	Source	**Year**	**Average**	**Std·Dev**	**Min**	**Max**
Demographic:						
-Density	ISTAT	2019	266·9	380·0	36 (Nuoro)	2574 (Napoli)
-Age	ISTAT	2019	45·85	1·62	41·67 (Napoli)	49·20 (Savona)
-Age index, percent	ISTAT	2019	195·4·	35·2	121·5 (Napoli)	275·8 (Biella)
-Mortality rate	ISTAT	2019	11·3	1·41	8·4 (Bolzano)	14·7 (Alessandria)
-Family size	ISTAT	2011	2·29	0·14	1·28 (Trieste)	3·45 (Napoli)
-Students, percent pop	ISTAT	2019	13·5	1·16	11·2 (Oristano)	16·5 (Napoli)
-Students in class, percent pop	ISTAT	2019	7·6	3·2	3·8 (Napoli)	16·5 (Ferrara)
-Share of secondary degree acquisition, percent 19+	ISTAT	2011	39.6	3.9	32.5 (Oristano)	54.2 (Roma)
-Share of postgraduate degree acquisition, percent pop	ISTAT	2011	1·71	0·47	0·59 (Trapani)	3·32 (Roma)
-Share Families 1 component	ISTAT	2011	31·07	4·19	20·11 (Barletta)	43·18 (Firenze)
-Share Families 5+ components	ISTAT	2011	5·72	1·95	2·46 (Trieste)	12·47 (Napoli)
Economics:						
-Income per capita, PPP, 10k euro	EUROSTAT	2017	39·6	3·93	32·95 (Oristano)	54·22 (Roma)
-Employment, percent pop	ISTAT	2019	38·9	6·3	25·7 (Crotone)	47·7 (Bolzano)
-Agriculture Share Population	ISTAT	2019	1·94	1·47	0·05 (Prato)	8·75 (Ragusa)
-Industry Share Population	ISTAT	2019	10·50	4·50	3·35 (Vibo V·)	19·62 (Belluno)
-Service Share Population	ISTAT	2019	26·50	4·41	17·28 (Caltanissetta)	37·84 (Roma)
-Retail and Accommodation	ISTAT	2019	8·19	1·49	5·06 (Caserta)	13·17 (Grosseto)
-Retail and Accommodation, open	ISTAT	2019	5·30	4·38	0 (Bergamo)	13·17 (Grosseto)
Commuting:						
-Work with public transport	ISTAT	2011	1·75	1·46	0·15 (Nuoro)	8·69 (Milano)
-Study with public transport	ISTAT	2011	3·47	0·78	1·23 (Sud Sardegna)	5·09 (Teramo)
-Concentration	ISTAT	2011	0·97	1·22	0·01 (Nuoro)	6·06 (Monza)
-Commuting covid IN	ISTAT	2011	0·24	0·20	0·01 (Palermo)	1·01 (Gorizia)
-Commuting covid OUT	ISTAT	2011	0·24	0·19	0·01 (Trapani)	0·85 (Vercelli)
-Commuting covid IN, public	ISTAT	2011	0·05	0·06	0·003 (Trapani)	0·34 (Trieste)
-Commuting covid OUT, public	ISTAT	2011	0·05	0·04	0·004 (Palermo)	0·33 (Gorizia)
Health:						
-Heart attack deaths per 1000 people	ISTAT	2019	2·20	0·42	1·28 (Sassari)	3·45 (Ferrara)
-Cancer deaths per 1000 people	ISTAT	2018	15·0	2·3	10·3 (Sassari)	20·18 (Alessandria)
-Increased life expectancy 2002-2017, years	ISTAT	2019	2·63	0·59	1·20 (Fermo)	4·60 (Gorizia)
-Asthma and COPD	Il Sole 24 Ore	2019	6·42	1·09	4·31 (Sud Sardegna)	9·65 (Benevento)
-Diabetes	ISTAT	2018	41·36	7·22	23·30 (Bolzano)	63·27 (Agrigento)
-Hypertension	Il Sole 24 Ore	2019	145·01	14·52	94·53 (Sud Sardegna)	186·40 (Ferrara)
-GPs per 1000 people	ISTAT	2019	0·93	0·16	0·52 (Nuoro)	1·38 (Rovigo)
-Hospital beds per per 1000 people	ISTAT	2017	3·41	0·88	1·55 (Sud Sardegna)	6·52 (Isernia)
Geograpichs:						
-Temperature 2007-2016	ISTAT	2016	15·35	1·76	11·43 (Belluno)	19·57 (Messina)
-First wave Covid incidence	Min. Salute	2020	24·46	23·31	1·80 (Sud Sardegna)	115·4 (Cremona)

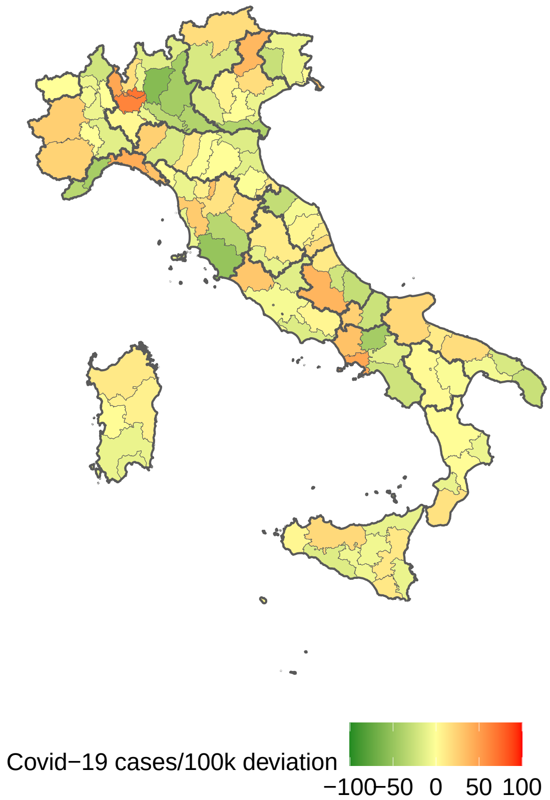
Note: The health data from il Sole 24 ore can be retrieved here: https://lab24.ilsole24ore.com/indice-della-salute/indexT.php

Appendix B Pre- and Post-Policy incidence

B.1 Pre-Policy incidence

Figure 3: **Weekly Cases per 100k people: Pre-Policy, 1/09/2020 - 3/11/2020**

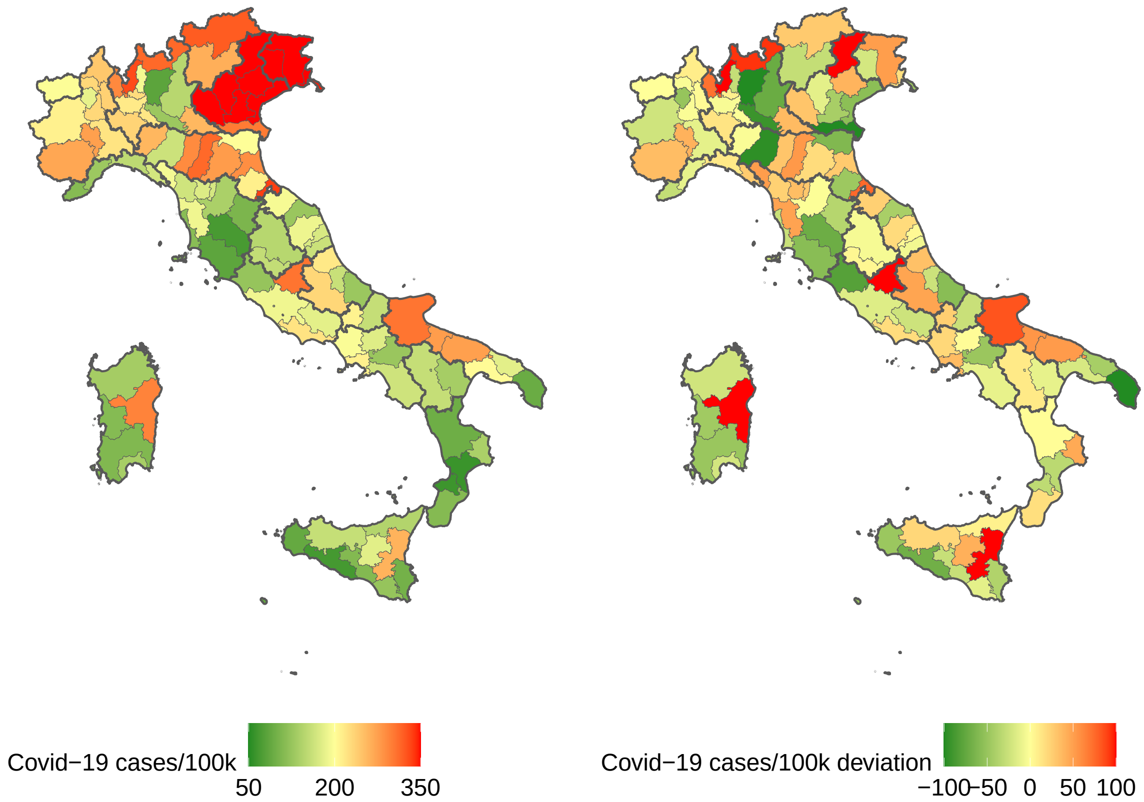
(a) Values



(b) Deviation

B.2 Post-Policy incidence

Figure 4: **Weekly Cases per 100k people: Post-Policy, 25/11/2020 - 23/12/2020**



(a) Values (b) Deviation

Appendix C Robustness tables

C.1 OLS estimates - Red and Yellow Tiers

Table 5: **Yellow and Red Tiers OLS Results**

	1st Sept. - 3rd Nov.				25th Nov. - 23rd Dec.			
	(1) Yellow Tier		(2) Red Tier		(3) Yellow Tier		(4) Red Tier	
Temperature	-2.122	(0.422)	-5.438**	(0.015)	-5.021	(0.483)	-24.31***	(0.000)
Income per Capita	2.675***	(0.004)	1.852*	(0.066)	-2.041	(0.407)	4.130	(0.115)
Agriculture Share Population	0.0956	(0.671)	-0.423**	(0.037)	-0.00667	(0.991)	-0.737	(0.161)
Services Share Population	0.433***	(0.008)	0.229**	(0.038)	0.679	(0.120)	0.516*	(0.073)
Share families 5+ components	5.983*	(0.083)	7.699***	(0.004)	15.89*	(0.089)	19.61***	(0.004)
Cases First Wave	-0.406***	(0.002)	-0.0683	(0.615)	-0.804**	(0.022)	-0.0195	(0.956)
Public Transport Trips Conc.	9.804***	(0.002)	9.746*	(0.069)	3.772	(0.652)	27.15*	(0.053)
Share Yellow Tier × Temperature	-2.690	(0.570)			-36.58***	(0.006)		
Share Yellow Tier × Income per Capita	-2.058	(0.300)			9.936*	(0.067)		
Share Yellow Tier × Agriculture Share Population	-0.903**	(0.027)			-0.334	(0.758)		
Share Yellow Tier × Services Share Population	-0.326	(0.185)			-0.172	(0.795)		
Share Yellow Tier × Share families 5+ components	1.784	(0.757)			8.391	(0.592)		
Share Yellow Tier × Cases First Wave	0.521**	(0.033)			0.862	(0.189)		
Share Yellow Tier × Public Transport Trips Conc.	-3.307	(0.702)			34.29	(0.147)		
Share Red Tier × Temperature			9.394	(0.105)			33.05**	(0.030)
Share Red Tier × Income per Capita			0.0316	(0.984)			-8.684**	(0.039)
Share Red Tier × Agriculture Share Population			0.847*	(0.066)			2.153*	(0.074)
Share Red Tier × Services Share Population			0.150	(0.608)			0.153	(0.841)
Share Red Tier × Share families 5+ components			-5.801	(0.438)			-11.24	(0.565)
Share Red Tier × Cases First Wave			-0.413**	(0.043)			-1.170**	(0.029)
Share Red Tier × Public Transport Trips Concentration			1.063	(0.871)			-23.69	(0.168)
Observations	104		104		104		104	
R^2	.773		.755		.833		.833	
$R^2(adj)$	.676		.65		.762		.761	
Region FE	Yes		Yes		Yes		Yes	
H_0	=(FE model)		See note		=(FE model)		See note	
F-Test	2.5 **		6.7 ***		3.9 ***		2 *	
Critical value (1% sign.)	2.9		2.9		2.9		2.9	

Note: Significance levels: * = 0.10; ** = 0.05; *** = 0.01. "Sh." stands for "Share. In the interaction terms, "Y" stand for "Yellow Tier" and "R" for "Red Tier". Number is parenthesis report the p-value of the t-test. All models are based on Equation 5. Specifications (1) and (3) test the model $\tilde{\beta}X_{ij} + \gamma^Y S_i^{\text{Yellow}} X_{ij}$ with null hypothesis $H_0:\ \tilde{\beta} + \gamma^Y = 0$. Specifications (2) and (4) test the model $\tilde{\beta}X_{ij} + \gamma^R S_i^{\text{Red}} X_{ij}$ against the null hypothesis $H_0:\ \tilde{\beta} + \gamma^R = 0$.

C.2 Robustness Checks

Table 6: **Robustness checks**

	25th Nov. - 23rd Dec.		4th Nov. - 26th Jan.	26/2/2020 - 26/1/2021
	(1) No Sardegna	(2) No SAR, CAM, SIC	(3) Extended	(4) All waves
Temperature	-17·88**	-20·11**	-10·34**	-1·261**
	(0·010)	(0·050)	(0·011)	(0·010)
Income per Capita	1·129	1·017	2·077***	0·680***
	(0·510)	(0·551)	(0·000)	(0·000)
Agriculture Share Population	-0·566	-0·339	-0·473*	-0·0638
	(0·220)	(0·592)	(0·062)	(0·244)
Services Share Population	0·506*	0·550*	0·372*	0·0715***
	(0·073)	(0·066)	(0·053)	(0·007)
Share families 5+ components	17·61***	22·39***	13·71***	1·827***
	(0·002)	(0·001)	(0·001)	(0·002)
Cases First Wave	-0·273	-0·288	-0·402***	0·200***
	(0·288)	(0·309)	(0·007)	(0·000)
Public Transport Trips Conc.	9·004	7·785	11·19***	2·719***
	(0·302)	(0·349)	(0·001)	(0·000)
Observations	99	85	104	104
R^2	·807	·815	·784	·914
$R^2(adj)$	·744	·75	·715	·886
Region FE	Yes	Yes	Yes	Yes
H_0	=(FE model)	=(FE model)	=(FE model)	=(FE model)
F-Test	3·5 ***	3·2 ***	7·1 ***	18·3 ***
Critical value (1% sign·)	2·9	2·9	2·9	2·9

Note: Significance levels: * = 0.10; ** = 0.05; *** = 0.01. All specifications use Conley Spatial Standard Errors with a cutoff of 150km. P-values of coefficients in parenthesis. "Public Tran. Trips Conc." stands for "Public Transport Trips Concentration". All regressions are controlled for region fixed effects. Therefore, the β coefficient on each variable can be interpreted as contributing to increasing (decreasing) Covid-19 cases per capita beyond (below) the regional mean. Specification (1) removers Sardegna due to its isolated status. Specification (2) removes also Campania and Sicilia, as they introduced some limited city-wide red tiers before the regional policies. Specification (3) extends the sample to 26th January 2021. Specification (4) considers the whole pandemic period.

Appendix D Robust Inference and Model Selection

The reader may be worried that the model selection through LASSO may change the inference approach that one should take in assessing the significance of the results. That is: can we really reject the null hypothesis that there are local-level effects in the pre-policy period, since we have selected the regressors in order to maximize R2 adjusted?

The worry here is that under small sample, the pre-selection over a large number of regressors may lead to overfitting and the selection of covariates uncorrelated to the dependent variable in the true data generating process, but correlated in the data due to small sample bias.

In this section, we show that simulating synthetic data allows us to produce an empirical distribution of post-selection OLS F-statistics under the null hypothesis. Using this distribution, we can build confidence intervals and rejection regions that account for the model selection algorithm. In particular, we generate 1000 draws of sets of 38 normally iid distributed regressors (random iid data, henceforth). We subtract regionals means in order to be centered within region. Then, we apply to each of them our model selection procedure and store the F-test p-value of the subsequent OLS regression (we take as reference specification 4, Table 1), assigning a value of one when no variable is selected ($\approx 15\%$ of the cases). Then, we check the 5th percentile of the distribution of p-values so obtained, which represents the critical value representing the OLS F-test p-value such that less than 5% of draws under the null hypothesis of no correlation between covariates and dependent variable sit at lower p-values. Finally, we compare this critical value with the p-value obtained in the real data. We repeat this exercise by drawing 1000 sets of 38 jointly normally distributed regressors, with covariance matrix replicating the one of our true dataset (random correlated data, henceforth). This allows to account for the preference of LASSO of selecting predictors with low correlation, selecting less variables than in the case of uncorrelated sets of regressors.

Our results are confirmed by this empirical, stricter rejection criteria, built to account jointly for the selection and post-selection steps. Table 7 shows how only 0.1% of the simulations in the iid data and 0% of the simulations in the correlated data have an F-test pvalue smaller than the one built using the real data. This is true whether we apply (right column) or do not apply (left column) the refinement process to maximize R2-adjusted after the LASSO. This means that the post-selection OLS p-value of the true data is much smaller than the one of most random data, with 99.9% of all simulations achieving a larger p-value. This means that our results are indeed significant at the 5% level and thus unlikely to be produced by covariates uncorrelated to the dependent variable.

In Table 8 we show similar results for the R2 adjusted: it is highly unlikely for randomly generated covariates to generate an amount of R2-adjusted similar to the one of the true data.

Table 7: **Random Generated Samples and Statistical significance, with and without refinement. Share of simulations**

	p-value(Fstat$_z$) $<$ p-value(Fstat Data)	
	Without refinement	With refinement
Random iid data	0.1%	0.1%
Random correlated data	0.0%	0.0%

Note: this table displays the share of simulations (out of 1000), in percent, for which the p-value of the F-statistics (null hypothesis: $H_0 : \tilde{\beta} = 0$, in model 2) is less than the one found in the data. The first row displays the results when the regressors are assumed to be iid. The second row displays the results when the regressors are assumed to have the same covariance matrix as the regressors in the data. The first column presents the results without the refinement, while the second column presents the results with the refinement.

Table 8: **Random Generated Samples and Explanatory power, with and without refinement: Additional R^2 Adjusted**

	Without Refinement		**With Refinement**	
	All Samples	Significant Samples	All Samples	Significant Samples
		Random iid data		
Average $R^{2,z}$	0.09	0.20	0.10	0.22
95% conf Interval	[0.0 - 0.21]	[0.16-0.25]	[0.0-0.22]	[0.18-0.26]
Frequency: $R^{2,z} > R^{2,data}$	0.3 %	6.0%	0.9%	14%
		Random correlated data		
Average $R^{2,z}$	0.07	0.18	0.08	0.21
95% conf Interval	[0.0 - 0.18]	[0.13-0.22]	[0.0-0.21]	[0.17-0.26]
Frequency: $R^{2,z} > R^{2,data}$	0.0%	0.0%	0.5%	6%

Note: this table displays the additional Adjusted R^2 of model 2 with respect to model 1 in the 1000 simulations. This statistic captures the additional explanatory power of the selected regressors in addition to the regional fixed effects. The top-panel displays the results when the regressors are assumed to be iid. The second panel displays the results when the regressors are assumed to have the same covariance matrix as the regressors in the data. The left panel presents the results without the refinement, while the right panel presents the results with the refinement. The first column presents the statistics for all the simulations (1000), while the second column presents the statistics for the 5% simulations with the lowest p-value of the F-statistics. The first line displays the average additional Adjusted R^2, across the simulations. The second line displays its 95 percent confidence interval. The third line displays the share of simulations, in percent, for which the Adjusted R^2 with the synthetic data is larger than the one found in the data (equal to 0.2449 without the refinement and equal to 0.2524 with the refinement).

Appendix E Post-LASSO Refinement Procedure

In this section, we discuss the role of the refinement to the LASSO selection discussed in the main text. The refinement works as follows: take all covariates selected by the LASSO procedure. Then, start iterating over the variables with the lowest p-value, perform an OLS regression and: (1) keep the variable if R2-adjusted does not increase, or (2) discard the variables if R2-adjusted increases. Under option (2), repeat the procedure until you find that R2-adjusted does not increase any further.

We have discussed in Appendix D how this has little impact on the inference procedure and on the explained R^2 adjusted of the selected model. In Table 9 we present further evidence of how the variable selection in random data and in a bootstrap exercise is affected by this refinement.[9] The refinement reduces the number of selected variables by 1.3 out of an average of 9.7 (when we use 38 random, uncorrelated regressors to simulate our procedure under the null hypothesis), and shrinks by 6 the upper bound of the 95% confidence interval of the distribution. When we simulate the procedure using correlated regressors with the same covariance matrix as the true data, the refinement shrinks the number of selected variables by 2 out of 8.2, and shrinks the upper bound of the confidence interval by 7 out of 29. Finally, when we bootstrap the error terms of the dependent variable, we find that the refinement shrinks the average selected covariates (from the true data) by 5.1 variables.

[9] In the bootstrap exercise we resample 1000 times the residual obtained after estimating Model 1 to recreate 1000 dependent variables that have the same systematic component as the one estimated from the data but a different realization of the random component. We then repeat all the steps of our methodology (Lasso selection, and refinement) to each newly obtained dependent variables.

Table 9: **Regressor Selection: with and without refinement**

	Without Refinement	**With Refinement**
	Random iid data	
Frequency 0 variables selected	13.9	13.9
Average Selected	9.7	8.4
95% conf Interval	[0 - 27]	[0-21]
	Random correlated data	
Frequency 0 variables selected	15.2	15.2
Average Selected	8.2	6.2
95% conf Interval	[0 - 29]	[0-22]
	Bootstrap	
Frequency 0 variables selected	0	0
Average Selected	21.4	16.3
95% conf Interval	[11 - 33]	[9-25]

Note: this table displays the share of simulations in which the selection procedure select zero regressors in percent, (first line); the average number of regressors selected (second line), and its 95% confidence interval (third line) obtained by using the Lasso procedure without (first column) and with (second column) our proposed refinement. The top and central panels display the results for the randomly generated data (iid and correlated, respectively). The bottom panel displays the results for the bootstrapping exercise.