# COVID-19 and non-communicable diseases: GMM/IV Panel VAR evidence from US states

**Abstract**

To analyse the endogenous connection between COVID-19 and non-communicable diseases (NCDs) in US states. I employ a panel VAR approach to investigate the endogenous interactions between deaths due to covid (COV) and NCDs in US states. The NCDs considered are respiratory (RES) and circulatory (CIR) diseases. I use Arellano–Bond's dynamic panel equations to supply estimates for the dynamic relationship between COVID-19 and NCDs. Thereafter, I estimate the orthogonalized impulse response functions (IRFs), which help isolate the response of a variable to an orthogonal shock in another variable of interest. The results show evidence of a significant positive relationship between mortality due to COVID-19 and NCDs. Most impacts peak after two weeks, corresponding with the latent period between covid infection and mortality. Besides, I find that while a positive shock to mortality from respiratory illnesses affects that from circulatory illnesses, the reverse is not the case. Lastly, these results differ by gender and are robust to several sensitivity checks, so large deviations are unexpected. The paper summarises the findings and proffers potential policies and interventions (e.g. providing cancer scanners in supermarkets) that could reduce the reinforcing effects of COVID-19 on NCDs.

**Keywords:** COVID-19; Mortality; NCDs; Panel VAR; U.S.

# 1 Introduction

COVID-19 (everywhere else, I shall simply refer to the disease as covid or coronavirus.) is the worst pandemic that has hit humankind since the last Century. Since its discovery in Wuhan, China, in December 2019 more than 500 million cases, with 6.32 million deaths have been recorded worldwide according to John Hopkins University as at 20th June 2022. Of this count, the U.S. leads in both the number of cases ($> 20\%$) and deaths ($> 18\%$). The novelty of this virus has necessitated a plethora of emerging research investigating its causes and consequences. Given the high infection rate of the virus, countries have initiated several containment measures to limit its spread while actively seeking for a permanent solution. These restrictions led to disruptions in several sectors of the economy, such as the health sector.

The disruption in health services following the pandemic resulted in routine health services being accessed less frequently, thus the ability of countries to address and respond to non-communicable diseases (NCDs) has been impacted. For example, new cancer diagnoses have fallen by 25% since the pandemic lockdown began in the Netherlands [1]. NCDs, especially cardiovascular diseases, cancers, diabetes, and chronic respiratory diseases, are the leading causes of death and disability globally and affect a larger proportion of people annually than all other causes combined - 71% of all deaths [2].

Medical studies suggest a potential link between covid and NCDs since those with pre-existing NCDs appear to be more vulnerable to contracting or even dying from the covid virus. A systematic review finds that patients with certain chronic illnesses such as diabetes, hypertension (and other cardiovascular diseases), chronic respiratory illnesses, chronic kidney and liver conditions are more likely to be affected by Covid-19 [3]. Similarly, another study use state-level Indian Covid-19 data to show a positive correlation between deaths per million with NCD risk factors such as obesity, hypertension, diabetes, *etc* [4]. They also show analogous findings using state-level Covid-19 cases per million. Some researchers use simple statistical analysis to describe the significant impact of the pandemic on NCD patients in Europe and the US [5]. They report that 50% of NCD patients report worsening conditions during the pandemic, while 17% developed a new

disease.

It is important to understand how NCD-related mortality has been impacted by the Covid pandemic in the U.S. since it has the largest share of infection worldwide. Given the lack of a detailed analysis in this area, this paper adds to the evidence by undertaking a more robust analysis that explores the bi-directional relationship between covid and NCDs in the US. While the existing few studies consider a subset of U.S. cities using raw death counts, this study takes a holistic approach by considering all covid-affected U.S. states. I carry out this activity by estimating panel-data vector autoregressions (PVARs) using weekly datasets covering all covid-affected states in the continental US for the period 2020-2021, allowing for substantial spatial and temporal variation in the model.

PVAR is increasingly becoming a popular econometric tool for estimating multivariate time-series data in a panel setting [6]. This is a class of models built on the achievements of traditional panel models [7], in terms of allowing the use of large dataset, and VAR models [8], that control for endogeneity by allowing endogenous interactions of the variables within the system. I use Arellano–Bond's dynamic panel equations to show and provide reliable estimates for the dynamic relationship between covid and NCDs. More importantly, this strategy allows the estimation of orthogonalized impulse response functions (IRFs), which helps isolate the response of a variable to an orthogonal shock in another variable of interest.

## 2   Methods

### 2.1   Methodology

I employ a panel VAR approach to investigate the endogenous interactions between deaths due to covid (COV) and NCDs in US states. NCDs are proxied here by respiratory (RES) and circulatory (CIR) diseases. The following model is estimated

$$y_{it} = \alpha_i + A(L)y_{it} + Bx_{it} + \varepsilon_{it} \tag{1}$$

where $y_{it}$ is a three-variable vector (COV, RES, CIR) in state $i$ at time $t$ (in weeks); $\alpha_i$ is a diagonal matrix of state-specific intercepts (fixed effects), which capture time-invariant factors that affect mortality (nation-wide lockdowns, for example). $A(L)$ is a matrix polynomial of lagged coefficients with $A(L) = A_1 L^1 + A_2 L^2 + \ldots + A_p L^p$, and $x_{it}$ is a vector of exogenous (weather) covariates as these have been cited as important influencers of mortality [9, 10]. I show in the robustness section that the results are sensitive to the exclusion of these exogenous controls. $B$ are parameters to be estimated, and $\varepsilon_{it}$ is a vector of idiosyncratic errors. The autoregressive order *(p=2)* of the VAR is selected using the Bayesian Information Criterion.

I transform death cases into their week-on-week (WoW) log-differenced values is to ease the interpretation of the impulse-responses in percentage terms and for policy relevance. I calculate the growth rate of mortality ($g_{it}$) as

$$g_{it} = log(y_{it}) - log(y_{it-1})$$

where $y_{it}$ refers to cumulative deaths from each mortality cause in state $i$ at time $t$. In principle, $y_{it} - y_{it-1}$ refers to the number of new deaths in the last one week within each state in the US. This decision follows from several papers [11]. Although, other studies [9] use a less lag period use a longer lag period to account for the mortality rates, there is evidence of no significant difference in the number of lags [12].

Given the dynamic nature of the model, the fixed effects are likely correlated to the lags of the dependent variable, meaning that the common method of eliminating fixed effects (mean-differencing), would produce biased results [13]. To overcome this weakness, I employ the forward mean-differencing or orthogonal deviation (Helmert transformation) approach proposed in [14] as an alternative transformation. According to [13], this "forward mean-differencing" approach removes the mean of all future observations for each state-week instead of using deviations from past observations. This transformation allows the use of lagged covariates as instruments since it retains the orthogonal structure between the lagged covariates and the transformed variables [15]. Hence, the model coefficients can be jointly estimated using system GMM.

To compute the impulse-response functions (IRFs) and forecast-error variance decompositions (FEVDs), it is important to apply Cholesky decomposition to the residuals to orthogonalize them. In the Cholesky ordering, I allow COV to have a contemporaneous impact on RES and CIR, while the latter two are not allowed to have such impact on the former. This arrangement, by construction, implies that the variable that appears earlier (covid) is weakly exogenous with respect to the rest of the covariates in the short run. However, with no strong theoretical basis for this ordering, I present results with alternative orderings as robustness.

Finally, the IRFs and FEVDs are estimated using the method described in [13], where the confidence intervals are estimated using Monte-Carlo simulations. These estimations were done using the *pvar* package in Stata by [16]. Practically, I re-estimated the IRFs by randomly building a draw of coefficients A of equation (1) using the estimated coefficients and the associated variance-covariance matrix. This process is repeated 1000 times to obtain the 5th and 95th percentiles of the distribution that is used as confidence intervals of IRFs. The IRFs describe the response of a variable over time to shocks to another variable within the system, while the variance decomposition measures the percent of the change in a variable that is explained by the innovation in another variable, at a given forecast horizon (12 weeks, in our case).
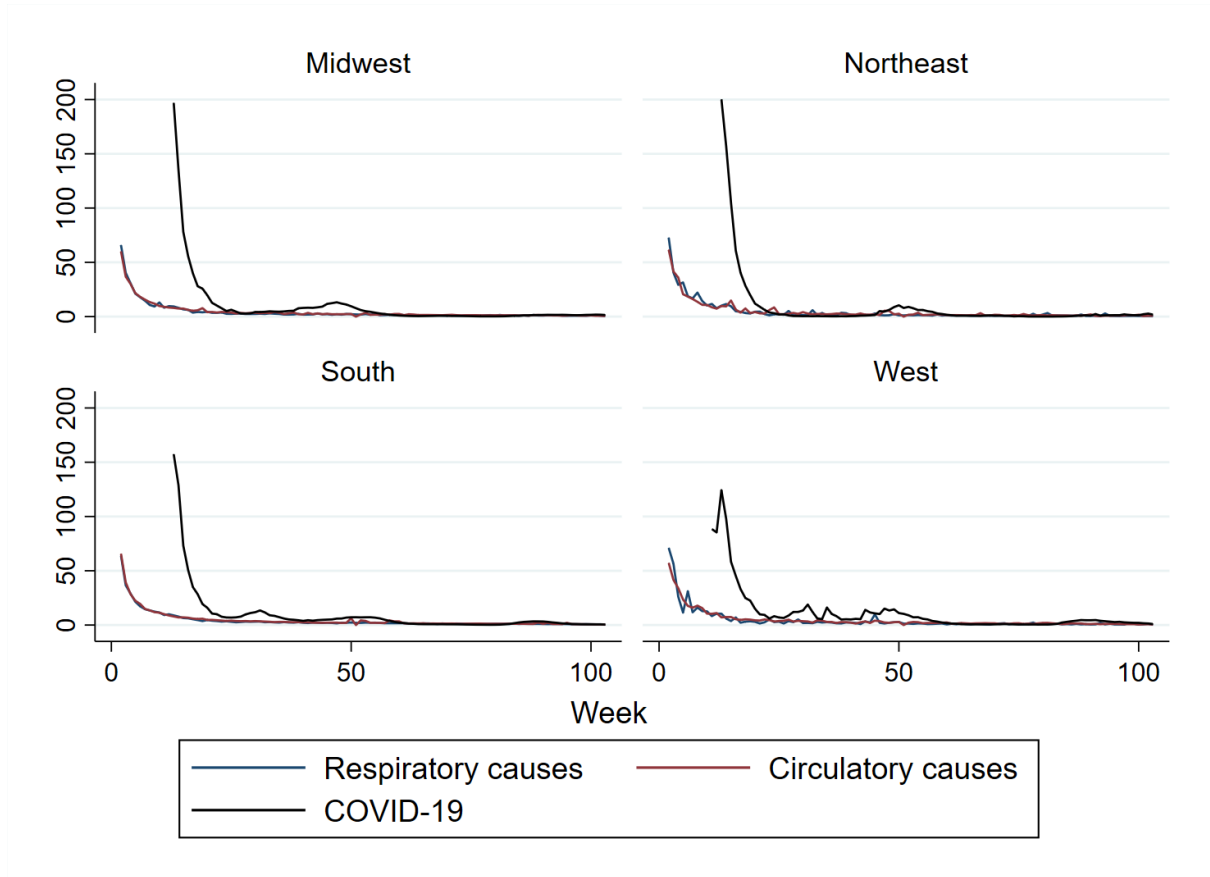
Figure 1: **Trends of growth rates of deaths (in %) from COVID-19, respiratory, and circulatory diseases in the U.S. (by Region)**

## 2.2    Data Sources and Description

**Mortality and Population Data**

I use weekly mortality data from three sources the United States Department of Health and Human Services (US DHHS), the Centers for Disease Control and Prevention (CDC), and the National Center for Health Statistics (NCHS). These data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program are freely available on the CDC WONDER Online Database. The dataset can be assessed *via* https://wonder.cdc.gov/wonder/help/mcd-provisional.html. More information on how the data are compiled can be found in the Technical Appendix domiciled on the website. Respiratory diseases are made up of 20 underlying causes of death related to the respiratory system, while circulatory diseases consist of 30 underlying causes of death related to the circulatory system. More information on how the data are compiled

can be found in the Technical Appendix domiciled on the website. The study's sample period begins from January 2020 to December 2021 to allow for the relevant period that captures the inception of Covid but excludes the vaccine era (2022). The sample contains the continental 48 U.S. states, including the District of Columbia, totalling 4,627 state-week observations. I excluded Alaska and Hawaii, and also dropped observations from unspecified states.

It is important to state that as at the time of collating the data, some of the observations were "provisional" implying they will likely not include all deaths that occurred during a given time period, especially for the more recent time periods. However, since this is a systemic issue and should not bias the results in any significant way. Further, state-level projected 2020 population information was obtained from the U.S. Census Bureau's website (`https://www.census.gov/topics/population.html`). I use this state-level population figure to derive the first difference in the natural logarithm of weekly cumulative deaths per 1 million people. Figure 1 suggests some comovement amongst the growth rates of mortality from the three diseases considered.

**Weather Data**

The historical weather dataset is obtained from the ERA5 reanalysis product from European Centre for medium-range weather Forecasts (ECMWF), which provides daily gridded weather variables at 0.25º resolution (see `https://cds.climate.copernicus.eu` for a complete description of the dataset). Specifically, I collected hourly downward UV radiation at the surface (in J/m²hour), 2-meter temperature (in ºC), 2-meter dewpoint temperature (in ºC), and 10 metre U and V wind components (in m/s) for January 01, 2020, to December 31, 2021. The weather measures were averaged across hours in a week to obtain weekly average measures. Thereafter, I link the weather data to state-level mortality cases by overlaying a US polygon with state boundaries on the gridded weather dataset for each grid cell and then taking a simple average across all grid cells per state.

# 3  Results and Discussion

Figure 2 depicts that estimated GMM panel VAR (Eq. (1)) is stable since the modulus of each eigenvalue of the fitted model is strictly less than one: inside the unit circle. The stability of the estimated model implies that shocks will eventually zero-converge: hence, the PVAR is invertible, thereby making the estimated IRFs and FEVDs interpretable.
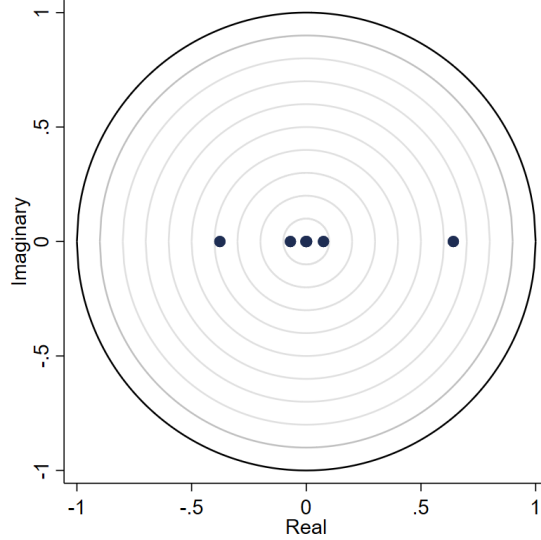


Figure 2: **Roots of companion matrix**

## 3.1  Main Results

In what follows, I present the IRFs graphs and their associated 95% confidence intervals generated *via* Monte Carlo simulations with 1,000 repetitions in Figure 3. The Figure reports the impact of a shock in any variable on other endogenous variables for 26 weeks (6 months) after the introduction of the shock. The results show that shocks to RES, CIR, and COV are transitory as the effect of a shock fades out almost immediately as seen in the diagonal panels (a, e, and i). The main interest in this paper, however, lies in the impact of shocks from any of the disease categories on the others – the off-diagonal panels. A positive shock to COV appears to have a positive positive impact on both RES and CIR as seen in panels g and h in Figure 3. Similarly, RES and CIR respective shocks also impact COV positively as seen in panels c and f. These results confirm that mortality from covid and NCDs are closely linked. Further, most impacts peak after

two weeks (which corresponds with the latent period between when one contracts covid and mortality) before they start dying out. Additionally, this study finds that a shock to CIR does not impact RES significantly, whereas a positive shock to RES affects CIR positively as seen in panels d and b, respectively. These results suggest that the perceived correlation between respiratory and circulatory disease-related deaths is due to the effect of RES on CIR, rather than the other way round.
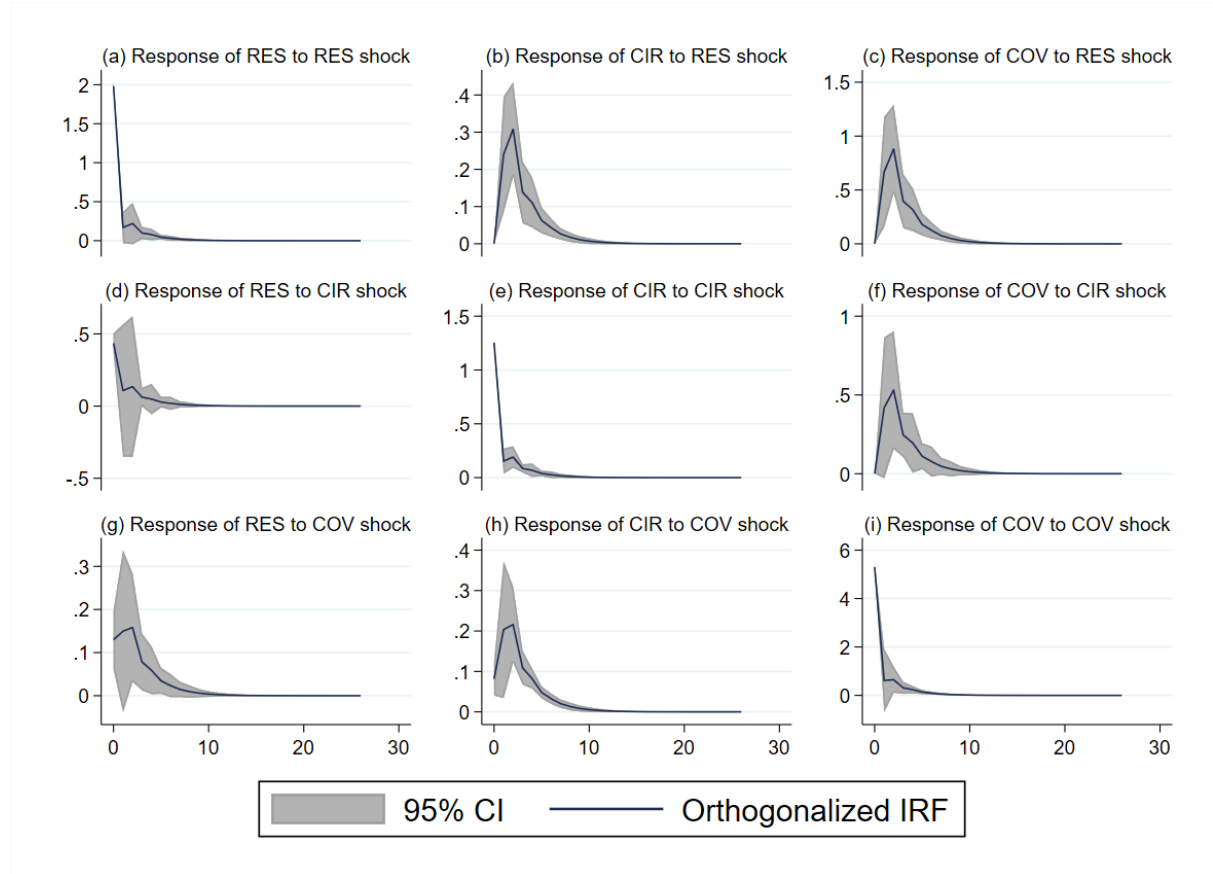


Figure 3: Impulse-response functions

I present the cumulative IRF in Figure 4 to show the effects in levels rather than in log-differences. This activity is done by accumulating the impacts over time (26 weeks). Although the appearance of Figure 4 differs from Figure 3, they have similar interpretations and findings. As earlier discussed, the impact of a positive shock to COV on RES and CIR is positive, and in like manner, a shock from either CIR or RES affects COV positively. Also, CIR shock does not have a significant impact on RES.
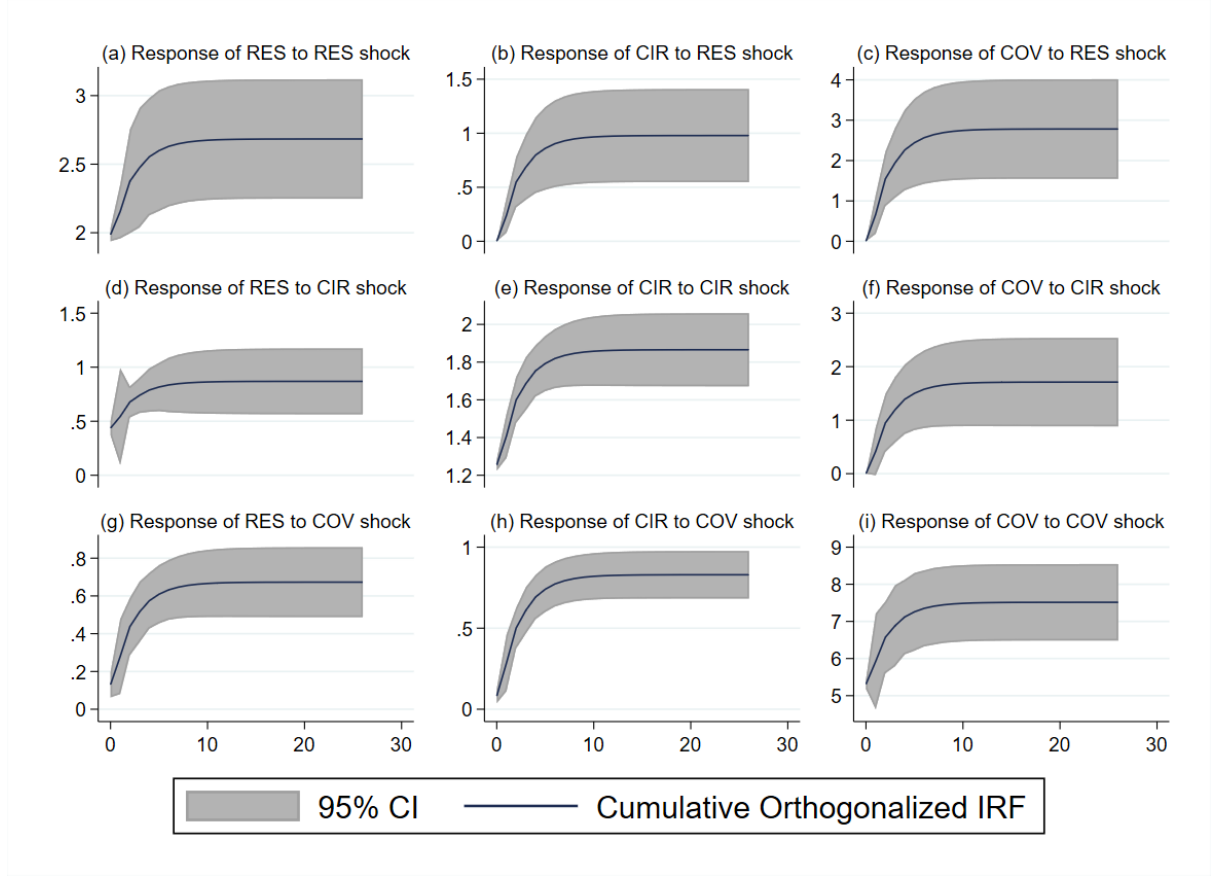
Figure 4: Cumulative IRFs

Advances in medical sciences have shown that mortality rate differs across sub-population [17]. Since the GMM approach imposes homogeneous dynamics across spatial units, I divide the sample by gender. The gendered IRFs in Figure 5 show that there exists heterogeneous dynamics across the sub-sample. While the overall pattern still remains qualitatively similar to the main results, the IRFs for both genders are below what is observed in the entire sample for responses to RES shock. On the other hand, the impact of COV shock is higher in males than females, suggesting that men with NCDs are more predisposed to die from covid than women.

Table 1: **Variance decompositions**

|       | COV   | CIR   | RES   |
| ----- | ----- | ----- | ----- |
| COV   | 0.932 | 0.018 | 0.049 |
| CIR   | 0.059 | 0.842 | 0.097 |
| RES   | 0.017 | 0.052 | 0.930 |

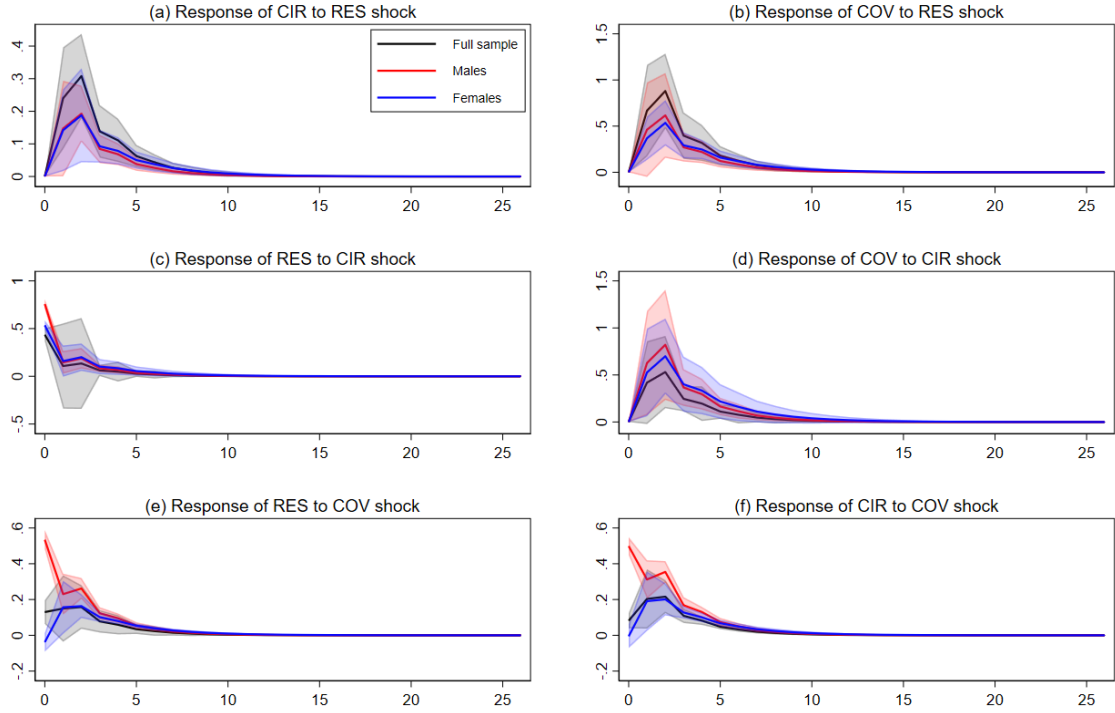*Note*: Variation in the row variable explained by column variable (25 periods ahead).



Figure 5: Gendered IRFs. 95% confidence bounds are represented by shaded areas.

To understand the contribution of changes in each variable to changes in other variables, I turn to the variance decomposition results presented in Table 1. The Table shows that RES explains more of the variation in COV and CIR for 25 periods ahead: nevertheless, the size of the effect is rather small, 4.9% and 9.7% respectively.

## 3.2 Robustness Results

Here, I use four alternative specifications to test the robustness of the results. First, I re-arrange the recursive order of the endogenous variables – allowing both RES and CIR

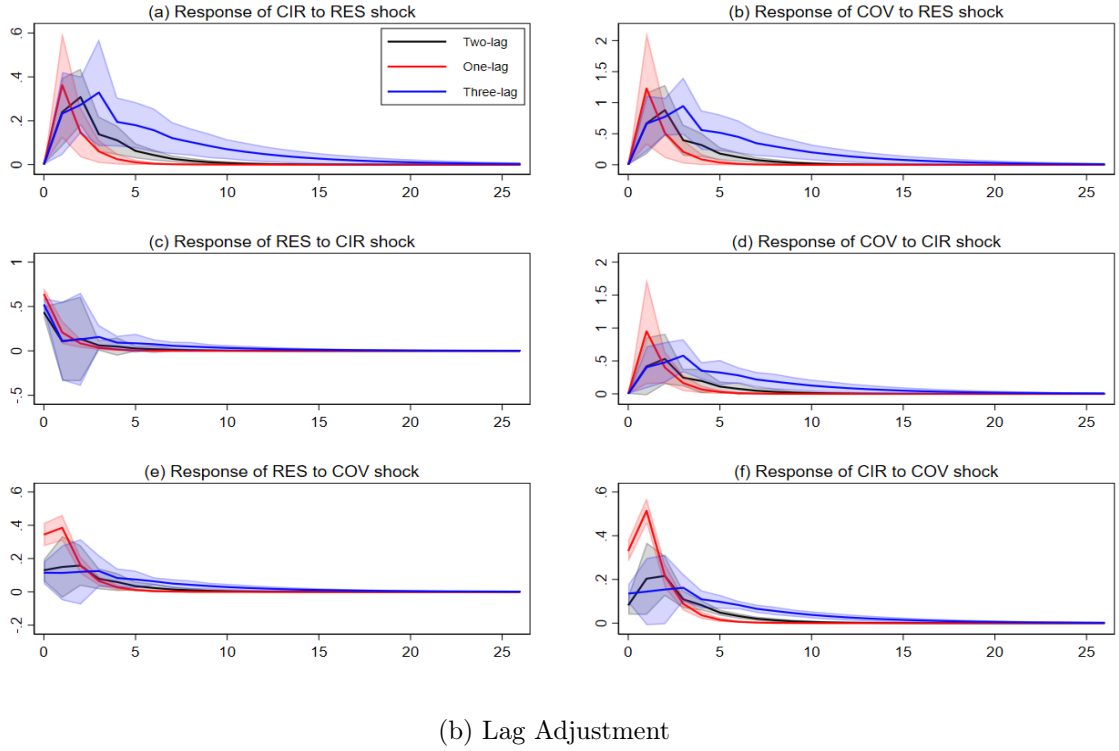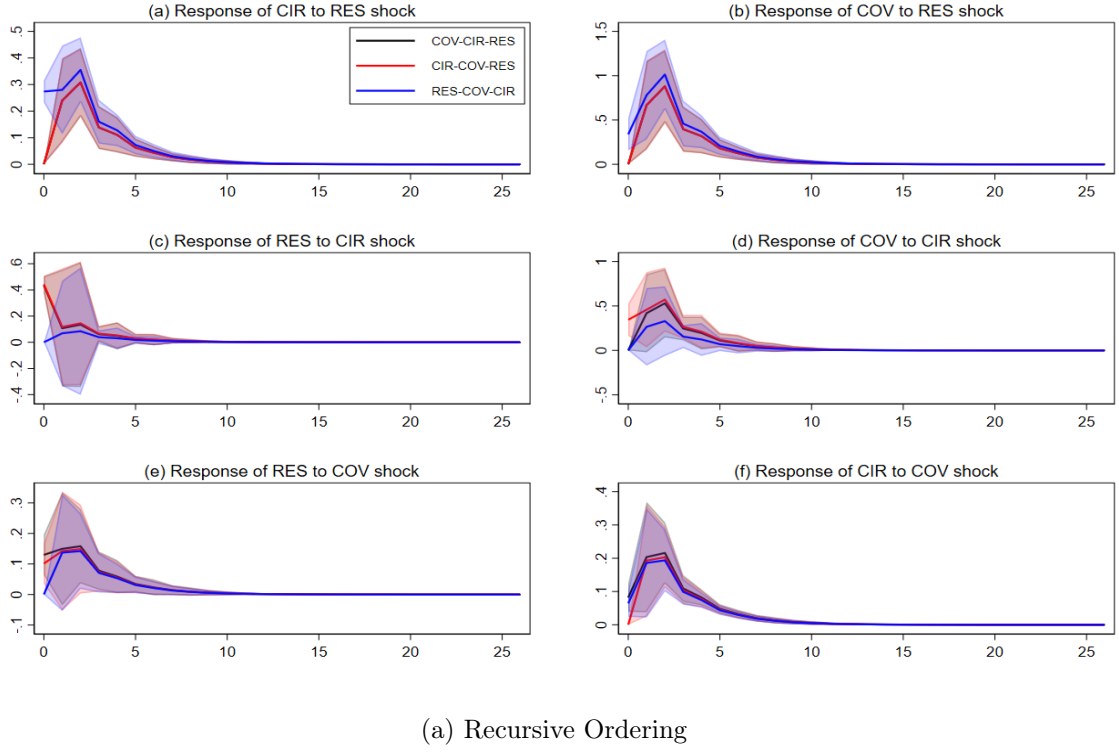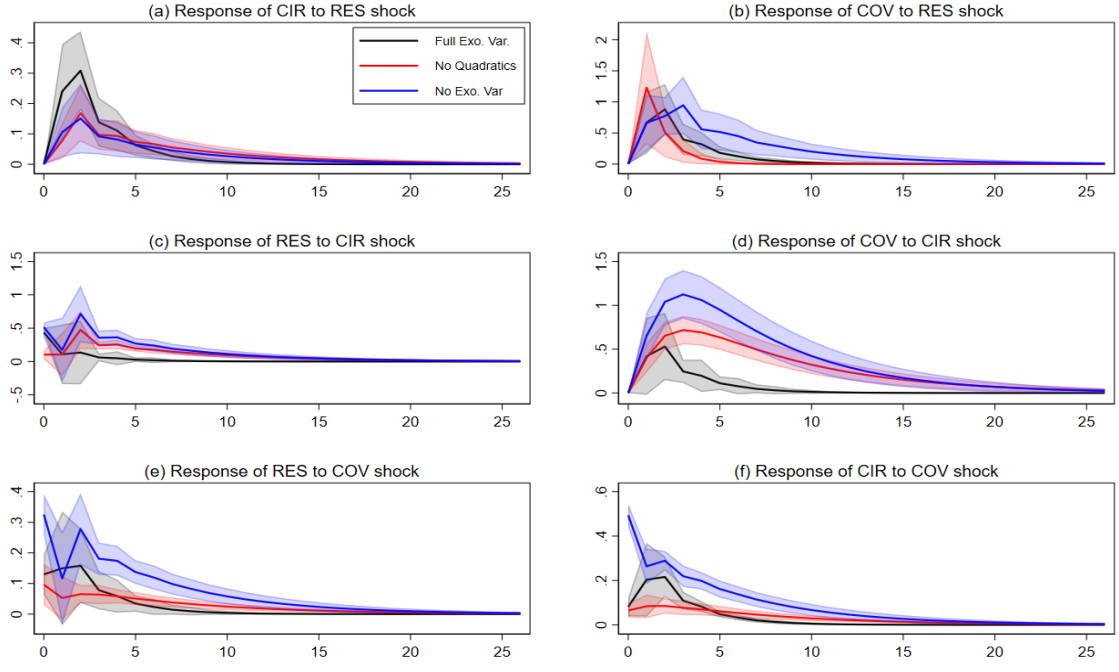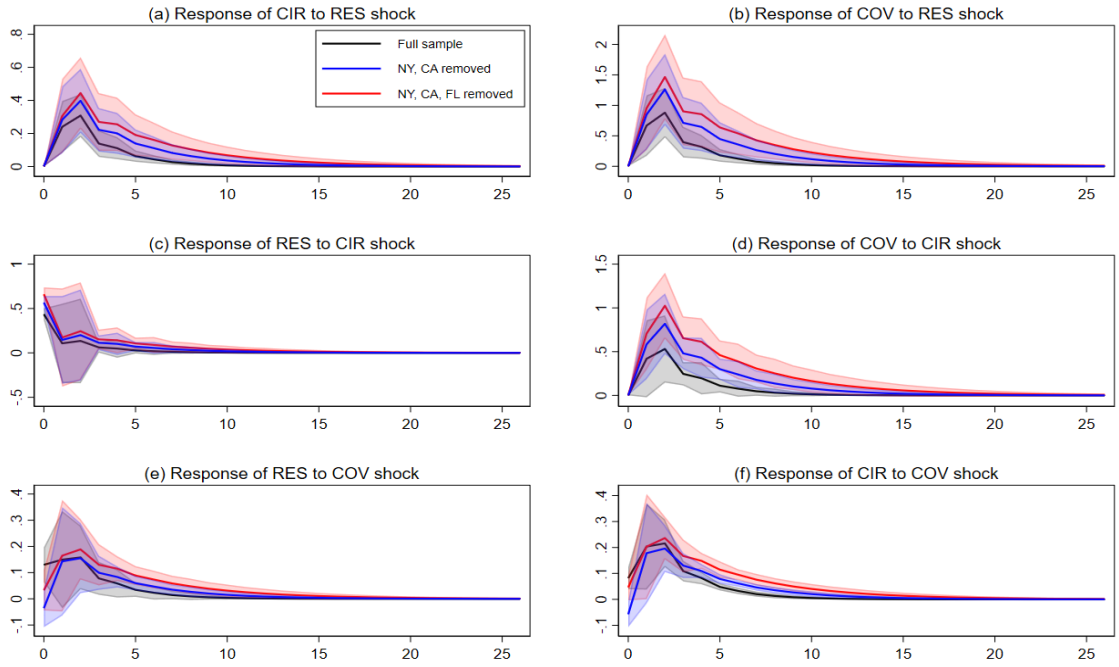(a) Recursive Ordering



(b) Lag Adjustment

Figure 6: IRFs from Alternative Specifications (Recursive Re-ordering and Lag Adjustment). 95% confidence bounds are represented by shaded areas.

(a) No exogenous variables



(b) Outliers

Figure 7: IRFs from Alternative Specifications (Weather Variables and Outliers). 95% confidence bounds are represented by shaded areas.

to assume weak exogeneity in turns. The results in Figure 6a produce a similar pattern as the main result. Also, I show in Figure 6b that the results are robust to lag choices by replacing the selected lag used in the primary model with lags 1 and 3, respectively. For most of the specifications, the IRF from 3-lag model appears higher than others, however, such small differences do not alter the overall interpretation of the results.

Next, the baseline model is re-estimated with the (weather) exogenous variables altered. Figure 7a shows that omitting weather variables significantly increases the impact of a shock on one variable on another variable. This result highlights the importance of weather variables in explaining mortality. Lastly, to confirm that the results are not driven by outliers, I purged the sample data of observation in the three states with the highest cumulative covid cases as of December 31 New York (NY), California (CA), and Florida (FL) in that order. Figure 7b shows that the states with large cases do not principally drive the results as the estimates are still very similar to the baseline results.

# 4   Conclusion

The results from this paper find that COVID-19 and NCDs reinforce themselves. After controlling for time-invariant location effects and important (exogenous) weather variables, I find a bi-directional causality between most combinations of death causes considered. Specifically, I find that covid mortality in the US positively affects mortality from respiratory and circulatory diseases. Although, the study does not find a statistically significant effect of mortality due to respiratory diseases on circulatory diseases. In addition, the paper finds that these impacts are higher in the male sub-population than in the females, which corroborates earlier studies [e.g., 18]. Finally, the results are robust to alternative VAR specifications and several samples, suggesting that large deviations are unexpected.

Using robust dataset and rigorous empirical approach, the findings in this paper add to evidence of the impact of coronavirus on NCD-related deaths. Urgent health policies, therefore, ought to be implemented to ensure that people dying from non-covid causes are not rising due to the enormous attention shifted towards the pandemic. Innovative health

measures (e.g., having cancer detection scanners in supermarkets) should be developed to ensure the continuation of conventional health services. Although with the introduction and push towards compulsory vaccination in countries like the US, it is expected that the coronavirus-NCD nexus will be weakened. However, further research is required to investigate if this assertion holds and to what degree.

# References

1. Dyer O. Covid-19: Pandemic is having "severe" impact on non-communicable disease care, WHO survey finds. 2020

2. WHO. Information note on COVID-19 and NCDs. 2020

3. Nikoloski Z, Alqunaibet AM, Alfawaz RA, Almudarra SS, Herbst CH, El-Saharty S, Alsukait R and Algwizani A. Covid-19 and non-communicable diseases: evidence from a systematic literature review. BMC Public Health 2021; 21:1–9

4. Gaur K, Khedar RS, Mangal K, Sharma AK, Dhamija RK and Gupta R. Macro-level association of COVID-19 with non-communicable disease risk factors in India. Diabetes & Metabolic Syndrome: Clinical Research & Reviews 2021; 15:343–50

5. Pécout C, Pain E, Chekroun M, Champeix C, Kulak C, Prieto R, Vugt J van, Gilchrist K and Lainé-Pellet AF. Impact of the COVID-19 Pandemic on Patients Affected by Non-Communicable Diseases in Europe and in the USA. International Journal of Environmental Research and Public Health 2021; 18:6697

6. Holtz-Eakin D, Newey W and Rosen HS. Estimating vector autoregressions with panel data. Econometrica: Journal of the econometric society 1988 :1371–95

7. Baltagi BH. Econometric analysis of panel data. John Wiley & Sons, London, 2001

8. Sims CA. Macroeconomics and reality. Econometrica: Journal of the Econometric Society 1980 :1–48

9.  Emediegwu LE. Health impacts of daily weather fluctuations: Empirical evidence from COVID-19 in US counties. Journal of Environmental Management 2021; 291:112662

10. Deschênes O and Greenstone M. Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. American Economic Journal: Applied Economics 2011; 3:152–85

11. Chernozhukov V, Kasahara H and Schrimpf P. Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. Journal of Econometrics 2021; 220. Themed Issue: Pandemic Econometrics / Covid Pandemics:23–62. DOI: `https://doi.org/10.1016/j.jeconom.2020.09.003`. Available from: `http://www.sciencedirect.com/science/article/pii/S0304407620303468`

12. Carleton T, Cornetet J, Huybers P, Meng KC and Proctor J. Global evidence for ultraviolet radiation decreasing COVID-19 growth rates. Proceedings of the National Academy of Sciences 2020; 118

13. Love I and Zicchino L. Financial development and dynamic investment behavior: Evidence from panel VAR. The Quarterly Review of Economics and Finance 2006; 46:190–210

14. Arellano M and Bover O. Another look at the instrumental variable estimation of error-components models. Journal of Econometrics 1995; 68:29–51

15. Baltagi BH. Econometric analysis of panel data. Vol. 4. Springer, 2008

16. Abrigo MR and Love I. Estimation of panel vector autoregression in Stata. The Stata Journal 2016; 16:778–804

17. Yang TC, Chen VYJ, Shoff C and Matthews SA. Using quantile regression to examine the effects of inequality across the mortality distribution in the US counties. Social Science & Medicine 2012; 74:1900–10

18. Aleksanyan Y and Weinman JP. Women, men and COVID-19. Social Science & Medicine 2022; 294:114698