GMM estimation of dynamic panel data models with invalid moment conditions—the sparse group lasso approach

Mingmei Xiao

Abstract

This paper primarily focuses on the GMM estimation of dynamic panel data models, where many moment conditions have been proposed under various assumptions. These moment conditions grow quadratically with the number of time periods T, making it difficult for researchers to determine which assumptions are satisfied in practice. Additionally, the presence of too many moment conditions can adversely affect the performance of the estimator. To address this, we explore the use of the sparse group lasso method for selecting valid moment conditions from a pool of potentially invalid ones. This paper reviews and compares existing methods with the sparse group lasso approach and provides simulation results to evaluate their performance. The application of the sparse group lasso method in dynamic panel data estimation demonstrates improved performance over the adaptive elastic net GMM approach proposed by Caner et al. (2018) in MSE, bias and standard errors. It performs similarly to the AL methods proposed by Andrews and Lu (2001), except in the most high-dimensional cases where AL performs better. However, AL is much more computationally costly than the Sparse group lasso approach. In addition, I apply this method to examine the effects of different Non-Pharmaceutical Interventions (NPIs) on mobility and how mobility influences the transmission of Covid-19. The results reveal that the effectiveness of policy measures varies significantly depending on the demographic characteristics of different areas, highlighting the need for more tailored policy approaches to effectively contain the spread of Covid-19.

Keywords: Sparse Group Lasso, Dynamic Panel Data Models, Covid-19

1 Introduction

Dynamic panel data models are useful and have proven to be popular in economic and social studies. They allow researchers to account for individual heterogeneity by controlling for unobserved time-invariant characteristics, instead of only focusing on aggregate time series behavior. Additionally, these models provide insights into the dynamic relationships within the data, which can be beneficial for impulse response studies. In this way, they are valuable for both causal inference and prediction. Over the years, dynamic panel data models (henceforth referred to as DPD models) have been applied to topics such as economic growth (Forbes, 2000; Levine et al., 2000), employment (Arellano and Bond, 1991; Blundell and Bond, 1998), and investment rates (Bond, 2002).

DPD models with fixed effects are known to suffer from the Nickell bias (Nickell, 1981) if estimated by demeaning or first differencing to remove the unobserved heterogeneity, as this transformation introduces correlation between the regressors and the error terms. Since then, several methods have been proposed for estimation using the Generalized Method of Moments (GMM), under various assumptions on the underlying data generating process, notably the Difference GMM (Arellano and Bond, 1991) and System GMM estimators (Blundell and Bond, 1998). Ahn and Schmidt (1995) proposed nonlinear moment conditions, which were recently found to be important for identification (Gørgens et al., 2019). However, two main concerns arise in empirical applications of these methods: the validity of moment conditions and the issue of instrument proliferation. Each set of moment conditions is based on a set of assumptions regarding the model specification, error terms, unobserved heterogeneity, and initial observations. The empirical problem at hand may or may not satisfy these assumptions, and most likely, the researcher doesn't have full knowledge of this. Moreover, the number of moment conditions proposed in the aforementioned papers grows quadratically with T, potentially leading to overfitting and poor estimation of the optimal weighting matrix, affecting efficiency and finite-sample performance (Roodman, 2009). Therefore, choosing the right model specifications and moment conditions is essential for consistent and efficient estimation of DPD models.

Regarding model and moment condition selection in GMM estimation, Andrews (1999) and Andrews and Lu (2001) introduced a well-known method. Their approach uses an information criterion-type penalized J-statistic to select the correct model specification and moment conditions. This method involves constructing several sets of possible models and moment conditions based on the assumed modeling assumptions. The criteria are then used to select the appropriate set of model and moment specifications for obtaining the estimates.

However, this approach is considered computationally expensive, especially when the set of potential combinations of models and moment specifications is large, as many repeated estimations and evaluations of the criteria are required for each combination.

This issue was recently addressed by Caner et al. (2018), who, instead of testing batches of moment conditions, included all possible variables and potential moment conditions in a general form during their GMM estimation. They used the adaptive elastic net method (Zou and Hastie, 2005) to simultaneously select both the model specification and the moment conditions. Their study was conducted under a general framework. However, when focusing on the dynamic panel framework, we can leverage our knowledge about the underlying assumptions behind the moment conditions and consider these conditions in groups, as in Andrews and Lu (2001).

In this paper, I aim to explore the use of the sparse group lasso method (Noah et al., 2013) for model and moment selection under various assumptions. This approach avoids testing different combinations of models and moments one at a time, as in Andrews and Lu (2001), while accounting for the fact that moment conditions in DPD models are often generated under the same set of assumptions and can thus be treated as valid or invalid together. This method differs from the group lasso method (Yuan and Lin, 2006) in that regularization is used to achieve both group-wise and individual-wise sparsity. This feature has two advantages in practical applications.

Firstly, when implementing model and moment selection simultaneously by incorporating both into the model—with moment conditions grouped into a few sets—the group lasso method may assign larger penalties to parameters related to moment selection than to those related to parameter selection, leading to suboptimal results. Even when penalties are adjusted, treating an entire group of moment conditions as either valid or invalid can be overly restrictive. Roodman (2009) provides an example where, if the conditions for initial observations in Blundell and Bond, 1998 are violated, the corresponding moment conditions for the early periods may be invalid, while the degree of violation diminishes over time, resulting in both invalid and locally valid moment conditions within the same group. In such cases, the sparse group lasso method offers greater flexibility by allowing for individual-level sparsity within groups.

In the simulation results, I find that the Sparse Group Lasso outperforms the adaptive elastic net approach proposed by Caner et al. (2018) in terms of bias, standard errors, and RMSEs. However, among all methods, the approach by Andrews and Lu (2001) yields the best performance, particularly in high-dimensional settings. Despite this, it is considerably more computationally expensive compared to the shrinkage methods.

Apart from theoretical interest, I also investigate how this method works when applied to the empirical question of the impact of mobility on the transmission of Covid-19 in the US. Researchers have proposed various methods for modeling Covid-19 since the start of the pandemic. However, Anthony Fauci, Director of the National Institute of Allergy and Infectious Diseases, has raised concerns about these models, famously stating, "Models are as good as the assumptions you put into them." The method I use in this paper helps relax assumptions within the DPD framework, and I apply this approach to model case transmission to provide better policy recommendations for managing Covid-19.

My approach is most similar to Wilson (2020), where they use rolling DPD to examine the sign and size of the impact and trace out the impact response paths for weeks ahead. Their results are questionable because they estimate the DPD model with fixed effects without correcting for the Nickell bias, which is particularly serious in a small T, large N setting. Simulations by Judson and Owen (1999) show that with $N = 100$ and T as large as 30, the bias may be as much as 20% of the true value of the parameter of interest. GMM methods developed so far could greatly improve the RMSE of the estimator. Hence, it is crucial to conduct the study using the newer methods. Since the Covid-19 situation in the US is constantly evolving, extending the period of analysis beyond the original study by Wilson (2020) is also important.

The paper is organized as follows. In Section 2, I set up the model and briefly discuss the existing methods in terms of their estimation performance. In Section 3, I formulate the sparse group lasso method. The following section presents the simulation study, comparing different models under various DGP settings. Then, the method is applied to an empirical problem to illustrate its usefulness.

2 Model and existing methods

2.1 Model and assumptions

The general model of interest is:

$$y_{it} = \alpha_1 y_{it-1} + \cdots + \alpha_p y_{it-p} + \beta^{\intercal} w_{it} + c_i + \tau_t + v_{it}, \tag{1}$$

where $t = p, \ldots, T$, and $i = 1, \ldots, N$. The random variables $\{y_{1it}\}$ are independent across individuals, have finite means, and satisfy the above relationship. The explanatory variables $\{w_{it}\} = (x_{1it}^\intercal, x_{2it}^\intercal, P_{1it}^\intercal, P_{2it}^\intercal, y_{2it}^\intercal)^\intercal$ include $\{x_{1it}\}$ and $\{x_{2it}\}$ as exogenous variables, $\{P_{1it}\}$ and $\{P_{2it}\}$ as predetermined variables, and $\{y_{2it}\}$ as endogenous variables. The $\{c_i\}$ are unobserved time-invariant individual effects, and $\{\tau_t\}$ are the time fixed effects shared between individuals. We can first eliminate the time-fixed effect by demeaning cross-sectionally to simplify our estimation. v_{it} is the error term associated with individual i at time t.

To estimate the parameters $(\alpha_1, \ldots, \alpha_p, \beta)$, we impose the following standard assumptions:

Assumption 1 (Standard Assumptions, SA):

$$E\left[v_{it}\right] = 0, \quad t = p, \ldots, T, \tag{2}$$

$$E\left[c_i\right] = 0, \quad \forall i, \tag{3}$$

$$E\left[c_i v_{it}\right] = 0, \quad \forall i, \quad t = p, \ldots, T, \tag{4}$$

$$E\left[v_{is} v_{it}\right] = 0, \quad \forall s \neq t, \quad \forall i, \tag{5}$$

$$E\left[v_{it} y_{i0}\right] = E\left[v_{it} y_{i1}\right] = \cdots = E\left[v_{it} y_{ip-1}\right] = 0, \quad \forall i, \quad t = p, \ldots, T. \tag{6}$$

Assumption 1 assumes that the individual effect and the error terms have zero means and are uncorrelated. The error terms are also serially uncorrelated and uncorrelated with the initial observations. Additional assumptions often made are:

Assumption 2 (Homoskedasticity):

$$E\left[v_{it}^2\right] = E\left[v_{ip}^2\right], \quad t = p, \ldots, T. \tag{7}$$

Assumption 3 ('Stationarity'):

$$E[y_{i0} c_i] = E[y_{i1} c_i] = \cdots = E[y_{ip} c_i]. \tag{8}$$

Assumption 2 ensures the error terms' variance remains constant over time. Assumption 3 asserts that the correlation between the individual effect and the initial observations remains constant.The failure of this assumption implies that the initial observations are not generated from the same distribution. It has been shown in Blundell and Bond (1998) to greatly improve the asymptotic efficiency of a GMM estimator of a dynamic panel data model with only one lag and no regressors, when the effect of the lagged variable is close to 1.

For the explanatory variables, we assume:

Assumption 4 (Explanatory Variables):

$$E\left[v_{it}\left(w_{ip}^{\intercal},\ldots,w_{it-1}^{\intercal},x_{1it}^{\intercal},x_{2it}^{\intercal},P_{1it}^{\intercal},P_{2it}^{\intercal}\right)^{\intercal}\right]=0,\quad t=p,\ldots,T, \tag{9}$$

$$E\left[v_{it}\left(x_{it+1}^{\intercal},\ldots,x_{iT}^{\intercal}\right)^{\intercal}\right]=0,\quad t=p,\ldots,T-1, \tag{10}$$

$$E\left[c_i\left(x_{1it}^{\intercal},P_{1it}^{\intercal}\right)^{\intercal}\right]=0,\quad t=p,\ldots,T, \tag{11}$$

$$E\left[c_i\left(x_{2it}^{\intercal},P_{2it}^{\intercal},y_{2it}^{\intercal}\right)^{\intercal}\right]=E\left[c_i\left(x_{2it-1}^{\intercal},P_{2it-1}^{\intercal},y_{2it-1}^{\intercal}\right)^{\intercal}\right],\quad t=p+1,\ldots,T. \tag{12}$$

The existing methods are mostly discussed in the AR(1) framework without explanatory variables:

$$y_{it}=\alpha y_{it-1}+u_{it}=\alpha y_{it-1}+c_i+v_{it}, \tag{13}$$

where $t=1,\ldots,T$ and $i=1,\ldots,N$. In the following sections, I will begin with this simpler model and later extend the discussion to the more general model presented earlier.

2.2 Moment Conditions

Many moment conditions have been proposed based on the above assumptions. The assumptions are written as a certain relationship between the observed variables and unknown parameters so that they could be passed into the GMM framework for estimation. The popular moment conditions that are used are the following:

From Assumption 1 (SA):

$$E\left[c_i\right]=E\left[v_{it}\right]=0\Rightarrow E\left[u_{it}\right]=E\left[y_{it}-\alpha y_{it-1}\right]=0,\quad t=1,\ldots,T\quad(\text{linear, } q=T). \tag{14}$$

$$E\left[c_iv_{it}\right]=E\left[v_{it}v_{is}\right]=0\Rightarrow E\left[u_{it}\Delta u_{it-1}\right]=0,\quad t=3,\ldots,T\quad(\text{nonlinear, } q=T-2). \tag{15}$$

$$\begin{aligned}&E\left[v_{it}y_{i0}\right]=0\quad(\text{along with } E\left[c_iv_{it}\right]=E\left[v_{it}v_{is}\right]=0)\\&\qquad\Rightarrow E\left[\Delta u_{it}y_{i0}\right]=E\left[\Delta u_{it}y_{i1}\right]=\cdots=E\left[\Delta u_{it}y_{it-2}\right]=0,\quad t=2,\ldots,T\\&\qquad(\text{linear, } q=\frac{(T-1)T}{2}).\end{aligned} \tag{16}$$

Additional moment conditions with Assumption 2:

$$
\begin{aligned}
&E\left[v_{it}^2\right] = E\left[v_{i1}^2\right] \quad (\text{along with } E\left[c_i v_{it}\right] = E\left[v_{it} v_{is}\right] = 0) \\
&\qquad \Rightarrow E\left[\left(u_{i1} + u_{i2} + \cdots + u_{iT}\right) \Delta u_{it}\right] = 0, \quad t = 2, \ldots, T \quad (\text{nonlinear, } q = T-1).
\end{aligned} \tag{17}
$$

Additional moment conditions with Assumption 3:

$$
\begin{aligned}
&E\left[y_{i0} c_i\right] = E\left[y_{i1} c_i\right] \quad (\text{along with } E\left[v_{it} y_{i0}\right] = E\left[c_i v_{it}\right] = E\left[v_{it} v_{is}\right] = 0) \\
&\qquad \Rightarrow E\left[u_{i2}\left(y_{i1} - y_{i0}\right)\right] = 0 \quad (\text{linear, } q = 1).
\end{aligned} \tag{18}
$$

The brackets after each condition indicates whether it is linear in the parameter to be estimated (i.e., α) and the number of moment conditions involved. Once expressed in terms of observed variables and unknown parameters, these conditions can be interpreted as restrictions on the first and second moments of the data. Equation 14 provides the only restriction on the first moments of the variable set $\{y_{1it}\}$. Equations 15 and 16, as shown by Ahn and Schmidt (1995), represent the moment conditions on the second moments of the data under Assumption 1 (SA). They demonstrate this by deriving the covariance matrix Λ for the observed variables $(u_{i1}, u_{i2}, \ldots, u_{iT}, y_{i0})$. After applying Assumption 1, the covariance matrix simplifies to:

$$
\Lambda = \begin{bmatrix}
\sigma_c^2 + \sigma_1^2 & \sigma_c^2 & \sigma_c^2 & \ldots & \sigma_{c0} \\
\sigma_c^2 & \sigma_c^2 + \sigma_2^2 & \sigma_c^2 & \ldots & \sigma_{c0} \\
\sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_3^2 & \ldots & \sigma_{c0} \\
\vdots & \vdots & \vdots & \ldots & \sigma_{c0} \\
\sigma_{c0} & \sigma_{c0} & \ldots & \sigma_{c0} & \sigma_{00}
\end{bmatrix} \tag{19}
$$

This is equally represented by the second moments of the observed data:

$$
\Lambda = E\begin{bmatrix}
u_1u_1 & u_1u_2 & u_1u_3 & u_1u_4 & u_1u_5 & \ldots & y_{i0}u_1 \\
u_2u_1 & u_2u_2 & u_2u_3 & u_2u_4 & u_2u_5 & \ldots & y_{i0}u_2 \\
u_3u_1 & u_3u_2 & u_3u_3 & u_3u_4 & u_3u_5 & \ldots & y_{i0}u_3 \\
u_4u_1 & u_4u_2 & u_4u_3 & u_4u_4 & u_4u_5 & \ldots & y_{i0}u_4 \\
u_5u_1 & u_5u_2 & u_5u_3 & u_5u_4 & u_5u_5 & \ldots & y_{i0}u_5 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
y_{i0}u_1 & y_{i0}u_2 & y_{i0}u_3 & y_{i0}u_4 & y_{i0}u_5 & \ldots & y_{i0}y_{i0}
\end{bmatrix}
$$

The matrix above shows how the moment conditions 15 and 16 completely restrict the pattern of the covariance matrix into the form implied by Assumption 1. The red equal signs represent condition 15 and the blue ones represent condition 16. To achieve the same pattern, we could have other variants of condition 15, such as:

$$E\left[u_{iT}\Delta u_{it}\right]=0, \quad t=2,\ldots,T-1, \tag{20}$$

or

$$E\left[u_{it}u_{it-1}\right]=E\left[u_{it}u_{it+1}\right], \quad t=2,\ldots,T-1, \tag{21}$$

which, when combined with 16, impose the same set of restrictions.

When we assume the homoskedasticity assumption, Ahn and Schmidt (1995) note that we can transform the nonlinear moment conditions 15 into linear ones. From 16:

$$E\left[\Delta u_{it}y_{it-2}\right]=0 \Rightarrow E\left[\Delta u_{it}y_{it-1}\right]=\alpha E\left[\Delta u_{it}y_{it-2}\right]+E\left[\Delta u_{it}u_{it-1}\right]=E\left[\Delta u_{it}u_{it-1}\right]. \tag{22}$$

Expanding the last term gives:

$$E\left[\Delta u_{it}u_{it-1}\right]=E\left[u_{it}u_{it-1}\right]-E\left[u_{it-1}^2\right]=E\left[u_{it-1}u_{it-2}\right]-E\left[u_{it-2}^2\right], \tag{23}$$

where the second variant of 15 and homoeskedasticity condition 17 are used for the second equality. Therefore we have

$$E\left[\Delta u_{it}y_{it-1}\right]=E\left[\Delta u_{it-1}y_{it-2}\right], \quad t=4,\ldots,T, \tag{24}$$

In this way, we transform the nonlinear moment conditions 15 to the above linear conditions.

When we assume the stationarity assumption, Blundell and Bond (1998) propose that 18, together with:

$$E\left[u_{it}\Delta y_{it-1}\right]=0, \quad t=3,\ldots,T, \tag{25}$$

makes the nonlinear moment conditions 15 redundant. To see this, note that under $E\left[v_{it}y_{i0}\right]=E\left[c_iv_{it}\right]=E\left[v_{it}v_{is}\right]=0$, we have:

$$E\left[y_{i0}c_i\right]=E\left[y_{i1}c_i\right] \Leftrightarrow E\left[u_{i2}\Delta y_{i1}\right]=0.$$

The above combined with Equation 25 imply $E\left[y_{it}c_i\right]=$ same for all t. Therefore:

$$E\left[\Delta y_{it-s}u_{it}\right]=0, \quad s=1,\ldots,t-1. \tag{26}$$

After imposing the above linear moment conditions, the nonlinear conditions 15 are automatically satisfied since

$$E\left[u_{it}\Delta y_{it-1}\right]=0\Rightarrow E\left[u_{it}\left(\alpha\Delta y_{it-2}+\Delta u_{it-1}\right)\right]=E\left[u_{it}\Delta u_{it-1}\right]=0. \tag{27}$$

Moreover, under both stationarity and homoskedasticity, using both

$$E\left[u_{it}\Delta y_{it-1}\right]=0,\quad t=3,\ldots,T, \tag{28}$$

and

$$E\left[\Delta u_{it}y_{it-1}\right]=E\left[\Delta u_{it-1}y_{it-2}\right],\quad t=4,\ldots,T, \tag{29}$$

in addition to the linear moment conditions in Assumption 3:

$$E\left[u_{i2}\Delta y_{i1}\right], \tag{30}$$

would make the nonlinear homoskedasticity moment conditions 17 and those in 15 redundant.

Generalizing the moment conditions to models with multiple lags and explanatory variables, as described at the start of this paper, is straightforward.

2.3 Moment and model selection methods

The problem I'm trying to solve arises when we know only that SA is satisfied, but we are unsure whether homogeneity, stationarity, or other sets of conditions hold. Several existing methods could be used to address this issue, and in this section, I provide a brief introduction to these methods.

AL's J statistics Andrews and Lu (2001) propose a method formulated on a very general dynamic panel data model, where the researcher is uncertain about the true lag length, the presence of trends, and the correlation structure between explanatory variables and the error term. Their objective function is based on the J-test for testing overidentification restrictions, with an additional penalty term that discourages the inclusion of too many variables and rewards the use of more moment conditions. They formulate their method using BIC, AIC, and HQIC criteria. These criteria are used to select the set of variables and moment conditions, after which post-selection GMM estimation is performed.Their simulation results show that estimators formed in this way exhibit lower biases, standard errors, root mean squared errors, and more accurate rejection rates than GMM without selection. In our context, their estimation procedure is as follows:

Given a particular selected subset of moment conditions (denoted by the selection vector c):

$$E\left[m_c\left(Y_i,\alpha\right)\right],$$

where Y_i represents the data $\{y_{it}\}_{t=0,\ldots,T}$ and α is a $p\times 1$ vector of parameters, let $G_c\left(Y_i,\alpha\right) = \frac{1}{n}\sum_{i=1}^{n} m_c\left(Y_i,\alpha\right)$, be their empirical estimates, and W_c be a weighting matrix with estimate $\widehat{W}_c\left(Y_i,\alpha\right)$. In GMM estimation, the optimal weighting matrix is:

$$\widehat{W}_c\left(Y_i,\alpha\right)^* = \left(\frac{1}{n}\sum_{i=1}^{n} m_c\left(Y_i,\alpha\right) m_c\left(Y_i,\alpha\right)^T\right)^{-1}.$$

The objective functions for moment and model choices are as follows:

$$MMSC_{\text{BIC}} = nG_c\left(Y_i,\alpha\right)^T \widehat{W}_c\left(Y_i,\alpha\right) G_c\left(Y_i,\alpha\right) - (|c|-|b|)\ln n, \tag{31}$$

$$MMSC_{\text{AIC}} = nG_c\left(Y_i,\alpha\right)^T \widehat{W}_c\left(Y_i,\alpha\right) G_c\left(Y_i,\alpha\right) - 2(|c|-|b|), \tag{32}$$

$$MMSC_{\text{HQIC}} = nG_c\left(Y_i,\alpha\right)^T \widehat{W}_c\left(Y_i,\alpha\right) G_c\left(Y_i,\alpha\right) - Q(|c|-|b|)\ln\ln n, Q>2 \tag{33}$$

where $|c|$ is the number of moment conditions selected and $|b|$ is number of nonzero model parameters selected. Subsets of moment conditions are evaluated using these objective functions, and the subset that minimizes the objective function will be selected. Then, the usual GMM estimation is applied to minimize:

$$G_c\left(Y_i,\alpha\right)^T \widehat{W}_c\left(Y_i,\alpha\right)^* G_c\left(Y_i,\alpha\right).$$

They also compare these methods with a downward testing procedure, where researchers start with the model and moment combinations that maximize the number of overidentifying restrictions, and step-by-step reduce these restrictions until they find a model that does not reject the null hypothesis that all the moment conditions are correct. Their simulations show that the MMSC-BIC and downward testing procedures work best, except for the smallest sample sizes.

The advantage of their method is its flexibility, as it can accommodate nonlinear moment and model choices. However, it is computationally expensive, even with the simpler downward testing procedure, because the optimization process needs to be repeated for each set of moment and model choices. If researchers aim to fully exploit this method by allowing for various features in dynamic panel data estimation, such as time-varying coefficients, structural breaks, and different correlation structures between explanatory variables, fixed effects, and error terms, the computational burden becomes significant.

Adaptive elastic net Caner et al. (2018) develop a method for GMM estimation that aims to achieve both model selection and moment selection by incorporating the largest model and all available moment conditions while penalizing large models and invalid moment conditions. Their framework can be adapted to fit the dynamic panel estimation context as well.

They first assign values to the moment conditions as follows:

$$E\left[m\left(Y_i,\alpha\right)\right]=F\tau, \tag{34}$$

where

$$F=\begin{bmatrix}0_{m-s,s}\\ I_s\end{bmatrix}, \tag{35}$$

with m moment conditions and s of them potentially invalid (the upper block corresponds to the valid conditions we are certain about, while the identity block represents the suspected conditions). τ is an $s\times 1$ vector (where only s_0 of them are truly nonzero due to invalid conditions).

They then form empirical estimates of $E\left[m\left(Y_i,\alpha\right)-F\tau\right]$ as:

$$G\left(Y,\theta\right)=\frac{1}{n}\sum_{i=1}^{n}\left(m\left(Y_i,\alpha\right)-F\tau\right), \tag{36}$$

where $\theta=\left[\alpha^T,\tau^T\right]^T$ is a vector of both model parameters and invalid moment conditions' deviation from zero.

Their adaptive elastic net GMM for the $(p+s)\times 1$ vector of parameters θ is defined as:

$$\widehat{\theta}=\left(1+\frac{\lambda_2}{\left(NT\right)^2}\right)\arg\min_{\theta}G\left(Y,\theta\right)^T\hat{W}G\left(Y,\theta\right)+\lambda_1\sum_{j=1}^{p+s}\widehat{\pi}_j\left|\theta_j\right|+\lambda_2\sum_{j=1}^{p+s}\theta_j^2, \tag{37}$$

where λ_1 and λ_2 are tuning parameters, $\hat{W}$ is a weighting matrix, typically chosen as:

$$\left(\frac{1}{n}\sum_{i=1}^{n}\left(m\left(Y_i,\alpha\right)-F\tau\right)\left(m\left(Y_i,\alpha\right)-F\tau\right)^T\right)^{-1}, \tag{38}$$

and p is the number of parameters to estimate (length of α). $\widehat{\pi}_j$ is the estimated weight on each coefficient. The term $\left(1+\frac{\lambda_2}{(NT)^2}\right)$ is included to reduce the bias caused by double shrinkage.

The estimation procedure is as follows: given a set of tuning parameters, we first estimate the elastic net estimator using the identity matrix as the weighting matrix:

$$\widehat{\theta}_{enet}=\arg\min_{\theta} G\left(Y,\theta\right)^T G\left(Y,\theta\right)+\lambda_1\sum_{j=1}^{s}|\theta_j|+\lambda_2\sum_{j=1}^{s}\theta_j^2. \tag{39}$$

From this step, we obtain $\widehat{\theta}_{enet}$. The weighting matrix $\hat{W}$ is then estimated as:

$$\left(\frac{1}{n}\sum_{i=1}^{n}\left(m\left(Y_i,\widehat{\alpha}_{enet}\right)-F\widehat{\tau}_{enet}\right)\left(m\left(Y_i,\widehat{\alpha}_{enet}\right)-F\widehat{\tau}_{enet}\right)^T\right)^{-1}, \tag{40}$$

and $\widehat{\pi}_j$ is estimated as $\left|\widehat{\theta}_j\right|^{-\gamma}$. In the second step, these are substituted into Equation 37 and minimizing the objective function would deliver the optimal model and moment parameters. The tuning parameters $\lambda=(\lambda_1,\lambda_2)$ are selected by minimizing the BIC-type information criterion:

$$G\left(Y,\widehat{\theta}\right)^T\hat{W}G\left(Y,\widehat{\theta}\right)+|S_\lambda|\ln(n)\max\left\{\ln(\ln(p+s)),1\right\}, \tag{41}$$

where $|S_\lambda|$ is the number of nonzero estimated coefficients under a tuning parameter choices λ.

Estimators developed under this setting show selection consistency and asymptotic normality, achieving the GMM efficiency bound (what can be achieved under full knowledge of the true moment conditions). Compared with previous literature, this method allows for moment conditions that hold only asymptotically to be selected as valid if the attached parameters decrease faster than a certain rate. The design of adding the l_2-norm ensures highly correlated variables remain in the model during selection. This way, correlated invalid instruments causing bias are all excluded, while correlated valid instruments improving efficiency are all retained.

Although this method was developed for GMM estimation, it was not specifically designed for dynamic panel data. In a brief section on its application to dynamic panel models, the only moment conditions they consider are Arellano and Bond (1991)'s linear moment conditions (equation 16 in our case) and the uncorrelatedness condition between the exogenous variable and the error term. The serial correlation assumption in their design is somewhat uncertain. In this paper, I aim to explore the dynamic panel setting more thoroughly, considering the violation of various assumptions. Additionally, the group characteristics of moment conditions under different assumptions are not utilized in their method, but this extra information could be used to improve the performance of the GMM estimator in dynamic panel data models.

The proposed sparse group lasso approach for model and moment selection in this paper can be extended to conduct GMM estimation, where the variables or moments have a grouping structure. Hence, the method also connects to the literature on using shrinkage methods for instrument selection. Kang et al. (2020) use lasso shrinkage for linear instrumental variable estimation. The intuition behind their method is that for instruments to be valid, the coefficient on them, when added to the second-stage regression, should be zero. This is derived from the moment condition that the instruments are uncorrelated with the second-stage error term. They also set up a GMM-type criterion function, with an ℓ_1 norm penalty term that penalizes the coefficients for the instruments, which is based on their identification condition that the number of invalid instruments must be below a certain threshold. Windmeijer et al. (2020) further extend this by using the adaptive lasso method to achieve better variable selection properties, resulting in improved results.

There is more literature on linear instrument selection using shrinkage methods, including Belloni et al. (2012), where they show that IV estimators using Lasso and Post-Lasso in the first stage are $\sqrt{n}$-consistent, asymptotically normal, and semi-parametrically efficient under homoskedasticity, assuming approximate sparsity. However, their method, like many others in this area, does not consider the invalid instrument scenario.

The method I develop in this paper is specifically designed for the dynamic panel setting, allowing for the grouping structure in moment conditions to be used, although it is currently constrained to linear models and moment conditions.

3 Selection using Sparse Group Lasso

This method extends the model used by Caner in Equation 37 in that the penalties on the model and moment parameters now take into account their grouping structure and performs group-wise and within-group selection. This is achieved by adopting the sparse group lasso estimator proposed in Noah et al., 2013. Continuing with previous notations, $E[m\left(Y_i, \alpha\right)] = 0$ is a stack of m moment conditions and can be written as:

$$E[m\left(Y_i, \alpha\right)] = E[B_i - A_i\alpha] = 0. \tag{42}$$

The sample moment conditions are therefore:

$$G\left(Y, \theta\right) = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(B_i - A_i\alpha - F\tau\right) = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(B_i - \begin{bmatrix} A_i & F \end{bmatrix}\begin{bmatrix} \alpha \\ \tau \end{bmatrix}\right), \tag{43}$$

which we can formulate as

$$G\left(Y, \theta\right) = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(B_i - X_i\theta\right), \tag{44}$$

where X_i is a $q \times (p + s)$ matrix. The sparse group lasso estimator for our dynamic panel model is defined as

$$\widehat{\theta}_{sgl} = \underset{\theta}{\operatorname{argmin}}\, G\left(Y, \theta\right)^T \hat{W} G\left(Y, \theta\right) + \lambda_1 \sum_{g=1}^{M} m_g \|\theta_{\varphi_g}\|_2 + \lambda_2 \|\theta\|_1 \tag{45}$$

where m_g is the weight for each group, φ_g indicates the coefficients in group g, and λ_1 is a tuning parameter. Typically, m_g is taken as $\sqrt{|\varphi_g|}$, which represents the cardinality of group g.

The properties of the Sparse Group Lasso are seen through the subgradient of its objective function: for each group k, $\hat{\beta}^k$ needs to satisfy the following for optimality:

$$\frac{1}{n}(X_k)^T(B - \sum_{l}^{M} X_l\hat{\beta}^l) = \lambda_1 m_g s_2 + \lambda_2 s_1, \tag{46}$$

where s_1 and s_2 are subgradients of the l_1 and l_2 norms evaluated at $\hat{\beta}^k$ and X_k is the subset of columns in X that correspond to group k parameters.

$$s_2 = \begin{cases} \frac{\hat{\beta}^k}{\|\hat{\beta}^k\|_2} & \text{if } \hat{\beta}^k \neq 0 \\ \in \{s_2 : \|s_2\|_2 \leq 1\} & \text{if } \hat{\beta}^k = 0 \end{cases} \tag{47}$$

$$s_{1j} = \begin{cases} \text{sign}(\hat{\beta}_j^k) & \text{if } \hat{\beta}_j^k \neq 0 \\ \in \{s_{1j} : |s_{1j}| \leq 1\} & \text{if } \hat{\beta}_j^k = 0 \end{cases} \tag{48}$$

Therefore, as discussed in Noah et al. (2013), for the whole group to be dropped, we need to satisfy a similar condition to Lasso, but with the coefficients in that group filtered through a soft shrinkage threshold controlled by λ_2 before applying the hard thresholding step. Within groups that are not dropped, the variables that are not selected satisfy the same conditions as in Lasso for a variable to be inactive, and the nonzero variables are estimated under elastic net-type conditions. Thus, overall, the shrinkage is more conservative at the group level than in the group lasso method and performs similarly to the elastic net at the individual level. This approach fits our purpose well, as within a group of moment conditions, some may be unsatisfied during the initial periods but become asymptotically satisfied later on.

Regarding computation, Noah et al. (2013) developed an algorithm that first checks if each group satisfies the condition to be dropped together. If not, individual estimates within each group are then solved using an accelerated generalized algorithm with backtracking. Since the penalties are separable between groups, their block gradient descent algorithm converges to the global minimum. In recent years, faster algorithms have been developed, such as the Fast Sparse Group Lasso Fujiwara et al. (2016), which improves speed by skipping the updates for groups where the parameters must be zero and focusing on updating groups where the coefficients must not be zero. This greatly improved our computational efficiency.

Another important consideration is that for optimal GMM estimation, we still need to calculate a first-step consistent covariance matrix for the moment conditions and use its inverse as the weighting matrix in the second-step estimation. In the simulations and empirical applications, we use the least restrictive GMM (the setting where we only use the certain moment conditions and the largest model) to form our first-step estimates and covariance matrix. To improve the accuracy of the inversion of this high-dimensional matrix (as in the sparse group lasso approach, we retain all potential moment conditions at the start, which grows quadratically with T), we apply linear shrinkage Ledoit and Wolf (2004).

4 Simulation study

For the simulation studies, I follow the setting of the empirical model in the subsequent section as closely as possible, where the researcher develops a model but faces uncertainty in terms of model specifications (lag length), as well as the underlying assumptions with their corresponding moment conditions. This setting mirrors that in the empirical section, where mobility is assumed to be predetermined but unsure whether it is strictly exogenous, as it could be influenced by past shocks to the growth rate of cases but remains uncorrelated with future shocks to the growth rate. Weather variables, on the other hand, are typically assumed to be strictly exogenous. I assume the model has one lagged effect, while the researcher is unsure whether there are 1 or 2 lags. This setting helps evaluate different methods' performances in selecting the appropriate lag length.

The true model

$$y_{it} = \alpha y_{it-1} + \beta_1 p_{it} + \beta_2 x_{it} + \eta_i + v_{it}, \tag{49}$$

where $t \geq 1$, y_{it} is the outcome variable, p_{it} is the predetermined but not strictly exogenous variable, x_{it} is the strictly exogenous variable, η_i is the unobserved time-invariant individual effect, and v_{it} is the error term. This is an AR(1) model with explanatory variables and satisfies the following assumptions: $\eta_i \sim N(0, \sigma_\eta^2)$, $v_{it} \sim N(0, \sigma_v^2)$, $E[\eta_i v_{it}] = 0\ \forall t$, $E[v_{is} v_{it}] = 0,\ \forall s \neq t$, and $E[v_{it} y_{i0}] = 0,\ t = 1, \ldots, T$. The variables from the true model are generated with the following correlation structure:

$$(x_{i1}, x_{i2}, \ldots, x_{iT}, p_{i1}, p_{i2}, \ldots, p_{iT}, \eta_i, v_{i1}, v_{i2}, \ldots, v_{iT})^T \sim N(0, \Sigma), \tag{50}$$

where

$$\Sigma = \begin{bmatrix} \sigma_x^2 I_T & \sigma_{xp} I_T & \sigma_{x\eta} 1_{T\times 1} & 0_T \\ \sigma_{xp} I_T & \sigma_p^2 I_T & \sigma_{p\eta} 1_{T\times 1} & \sigma_{pv} P \\ \sigma_{x\eta} 1_{1\times T} & \sigma_{p\eta} 1_{1\times T} & \sigma_\eta^2 & 0_{1\times T} \\ 0_T & \sigma_{pv} P^T & 0_{T\times 1} & \sigma_v^2 H. \end{bmatrix} \tag{51}$$

$H = \begin{bmatrix} h_1 & 0 & 0 & \dots & 0 \\ 0 & h_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & h_T \end{bmatrix}$ leads to homoskedasticity if $h_1 = h_2 = \cdots = h_T$, while

$$P = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \tag{52}$$

indicates that the predetermined variable is only correlated with the error term from the last period.

Initial Observation and Stationarity The generation of the initial observation takes into account the 'stationarity' assumption:

$$E[\eta_i y_{i0}] = E[\eta_i y_{i1}] \Rightarrow E[\eta_i y_{i0}] = \alpha E[\eta_i y_{i0}] + \beta_1 \sigma_{p\eta} + \beta_2 \sigma_{x\eta} + \sigma_\eta^2, \tag{53}$$

which leads to:

$$E[\eta_i y_{i0}] = \frac{\beta_1 \sigma_{p\eta} + \beta_2 \sigma_{x\eta} + \sigma_\eta^2}{1-\alpha}. \tag{54}$$

Therefore, the initial observation is generated as:

$$y_{i0} = \frac{\beta_1 \sigma_{p\eta} + \beta_2 \sigma_{x\eta} + \sigma_\eta^2}{(1-\alpha)\sigma_\eta^2} \phi \eta_i + v_{i0}, \tag{55}$$

where $\phi = 1$ and $v_{i0} \sim N(0, \sigma_v^2)$, independent from the other error terms, assuming the stationarity condition is satisfied.

Parameter settings The parameter settings are taken from a first-stage GMM estimation of the relationship between mobility and Covid case growth, assuming that it follows the true model. A daily panel of US counties is used for this exercise. Specifically, the model we estimated in generating the parameter setting is:

$$\Delta case_{it} = \alpha \Delta case_{i,t-1} + \beta_1 \Delta mob_{i,t-1} + \beta_2 \Delta test_{i,t} + \alpha_i + \epsilon_{it}, \tag{56}$$

where Δcase_t is the 100*(log difference of cumulative cases) for period t, Δmob_{t-1} is the first-differenced mobility index MEI (detailed in Section 5.1), and $ltest$ is 100*(log difference of cumulative tests). All the variables are first demeaned cross-sectionally to eliminate time fixed effects. Lags of orders 5 to 10 of the dependent variable are used to estimate the model with GMM. Both the Sargan's J test and the autocorrelation tests are passed. The estimation result is shown in Table 1. Once we have the parameter estimates, we can recover the individual fixed effects by obtaining the fitted values for the model as well as the aggregate residual (fixed effect + residual). Averaging each county's residuals over time gives the county fixed effects, and subtracting the fixed effects from the aggregated residuals gives the residuals. Then, we can calculate the variances and covariances between these variables and other regressors in the model. In the end, we obtain the following set of parameter values:

$$(\alpha, \beta_1, \beta_2, \sigma_x^2, \sigma_p^2, \sigma_\eta^2, \sigma_v^2, \sigma_{xp}, \sigma_{x\eta}, \sigma_{p\eta}, \sigma_{pv}) = (0.7, 0.07, 0.05, 4000, 50, 2.7, 60, 24, 0.43, -0.43, -11). \tag{57}$$

Table 1: Preliminary regression on parameter values

	Dependent variable: $\Delta case_t$
	$\Delta case_t$
$\Delta case_{t-1}$	0.708***
	(0.034)
Δmob_{t-1}	0.069**
	(0.031)
$\Delta test_t$	0.051**
	(0.023)
Sargan's J Test P-value	0.438
Observations	879

*p<0.1; **p<0.05; ***p<0.01

The $h_1, h_2, \ldots, h_T$ values that characterize the heteroskedasticity pattern are generated arbitrarily, with half of the values taking 1 and the other half taking 2. The sample sizes I consider are $(T, N) = (10, 500), (10, 1000), (30, 500), (30, 1000)$, where the last case is experimented due to its resemblance to the empirical setting. For each setting, I perform 1000 iterations.

The researcher's model

$$y_{it} = \alpha_1 y_{it-1} + \alpha_2 y_{it-2} + \beta_1 p_{it} + \beta_2 x_{it} + \eta_i + v_{it} \tag{58}$$

where $t = 2, \ldots, T$. The researcher knows the "SA" assumptions are satisfied but is unsure about the homogeneity and 'stationarity' assumptions. Regarding the explanatory variables, the researcher knows that x_{it} is strictly exogenous but is uncertain whether p_{it} is predetermined or strictly exogenous. The pool of linear moment conditions to use includes: Conditions known to be valid (*1st set*):

$$E\left[y_{it} - \alpha_1 y_{it-1} - \alpha_2 y_{it-2} - \beta_1 p_{it} - \beta_2 x_{it}\right] = 0, \quad t = 2, \ldots, T$$

The length of moment conditions above is $T-1$.

$$\begin{aligned} E\big[(y_{i0}, y_{i1}, \ldots, y_{it-2})\big(\Delta y_{it} - \alpha_1 \Delta y_{it-1} - \alpha_2 \Delta y_{it-2} \\ - \beta_1 \Delta p_{it} - \beta_2 \Delta x_{it}\big)\big] = 0, \quad t = 3, \ldots, T \end{aligned}$$

The length of moment conditions above is $\frac{(T+1)(T-2)}{2}$.

$$\begin{aligned} E\big[(x_{i1}, x_{i2}, \ldots, x_{iT})\big(\Delta y_{it} - \alpha_1 \Delta y_{it-1} - \alpha_2 \Delta y_{it-2} \\ - \beta_1 \Delta p_{it} - \beta_2 \Delta x_{it}\big)\big] = 0, \quad t = 3, \ldots, T \end{aligned}$$

The length of moment conditions above is $T(T-2)$.

$$\begin{aligned} E\Big[(p_{i1}, \ldots, p_{it-1})\Big(\Delta y_{it} - \alpha_1 \Delta y_{it-1} - \alpha_2 \Delta y_{it-2} \\ - \beta_1 \Delta p_{it} - \beta_2 \Delta x_{it}\Big)\Big] = 0, \quad t = 3, \ldots, T \end{aligned}$$

The length of moment conditions above is $\frac{(T+1)(T-2)}{2}$.
Additional conditions are valid if homoskedasticity holds (*2nd set*):

$$\begin{aligned} &E\Big[y_{it-1}\Big(\Delta y_{it} - \alpha_1 \Delta y_{it-1} - \alpha_2 \Delta y_{it-2} - \beta_1 \Delta p_{it} - \beta_2 \Delta x_{it}\Big)\Big] \\ &\quad = E\Big[y_{it}\Big(\Delta y_{it+1} - \alpha_1 \Delta y_{it} - \alpha_2 \Delta y_{it-1} - \beta_1 \Delta p_{it+1} - \beta_2 \Delta x_{it+1}\Big)\Big], \quad t = 3, \ldots, T-1 \end{aligned}$$

$$
\begin{aligned}
\Rightarrow E\Big[&(y_{it-1}\Delta y_{it} - y_{it}\Delta y_{it+1}) - \alpha_1(y_{it-1}\Delta y_{it-1} - y_{it}\Delta y_{it}) \\
&- \alpha_2(y_{it-1}\Delta y_{it-2} - y_{it}\Delta y_{it-1}) - \beta_1(y_{it-1}\Delta p_{it} - y_{it}\Delta p_{it+1}) \\
&- \beta_2(y_{it-1}\Delta x_{it} - y_{it}\Delta x_{it+1})\Big] = 0, \\
&\quad t = 3, \ldots, T-1
\end{aligned} \tag{59}
$$

The length of moment conditions above is $T-3$.
Additional conditions are valid if stationarity holds (*3rd set*):

$$
E\Big[\Big(y_{it} - \alpha_1 y_{it-1} - \alpha_2 y_{it-2} - \beta_1 p_{it} - \beta_2 x_{it}\Big)\Delta y_{it-1}\Big] = 0, \quad t = 2, \ldots, T \tag{60}
$$

The length of moment conditions above is $T-1$.
If both homoskedasticity and stationarity conditions are satisfied, we would have both sets of moment conditions above (*2nd* and *3rd* sets).
Additional conditions if p_{it} is strictly exogenous (*4th set*):

$$
\begin{aligned}
E\Big[(p_{it}, p_{it+1}, \ldots, p_{iT})\Big(&\Delta y_{it} - \alpha_1 \Delta y_{it-1} - \alpha_2 \Delta y_{it-2} \\
&- \beta_1 \Delta p_{it} - \beta_2 \Delta x_{it}\Big)\Big] = 0, \quad t = 3, \ldots, T
\end{aligned}
$$

The length of moment conditions above is $\frac{(T-2)(T-1)}{2}$.
Now we have all 4 sets of moment conditions the researcher may think of using. In terms of model specification, the researcher has 3 choices: lag length of 1, or 2.

Under the correct model, α_2 should be 0, and with different settings of the correct model, the corresponding correct moment conditions to use are the following:

non-"stationary" and heteroskedastic	Set 1
"stationary" and heteroskedastic	Set 1, 3
non-"stationary" and homoskedastic	Set 1, 2
"stationary" and homoskedastic	Set 1, 2, 3

Performance measurements I compare the results for the following methods: Andrews and Lu (2001)'s moment and model selection by information criteria (including AL-AIC, AL-BIC, AL-HQIC), Caner et al. (2018)'s method of selection by adaptive elastic net (Caner), the group lasso approaches proposed in this paper, benchmark GMM estimators using the correct model (Correct), the biggest model with all moment conditions (Biggest), the smallest model with minimum moment restrictions (Smallest), the most restrictive model with the smallest model and all moment conditions (Most), and the least restrictive model with the biggest model and minimum moment conditions (Least). Note that the GMM estimation for all methods are two-step GMM with the optimal weighting matrix calculated as the inverse of the first-step covariance of the moment conditions.

The performance measurements include the averaged bias, standard error, and RMSE of the estimators for the parameters in our model.

Simulation results The simulation results are reported as bar plots in Figure 1, 2, 3, and 4. These figures show for difference performance metric and sample size combinations, the mean value across all the parameters in our model, with each color representing an estimation method. The results for the 4 different settings on whether the model satisfies homoskedasticity/'stationarity' seem similar. AL methods perform the best out of the candidate methods but the computation time greatly scales up when we have more candidate model and moment choices. The Sparse group lasso method shares similar performance as AL methods in most scenarios but with much faster speed. The most noticeable exception where Sparse group lasso performs significantly worse compared with the AL method is for when $T = 30$ and $N = 500$, the most high-dimensional setting in our simulation. Caner's adaptive elastic net method also performs badly in this scenario. This is largely related to the bias introduced by the shrinkage methods, motivating further work to be done on better bias-corrected versions of these estimators. Compared with Caner's adaptive elastic net, our sparse group lasso method generally performs better in terms of MSE, bias, and variance. We also observe that the biggest and the least restrictive models work significantly worse than the others. Given our true model is sparser with only 1 lag, this highlights how getting the model choices wrong can affect the results way more than getting the moment choices wrong, as the effect of the latter can be diluted by other correct moment conditions. For all methods, when we increase N, for a given T, the MSE, bias, and variance tend to decrease as empirical moments are estimated more accurately. When we increase T, for a given N, MSE and bias tend to increase whereas the parameter's variances tend to decrease.

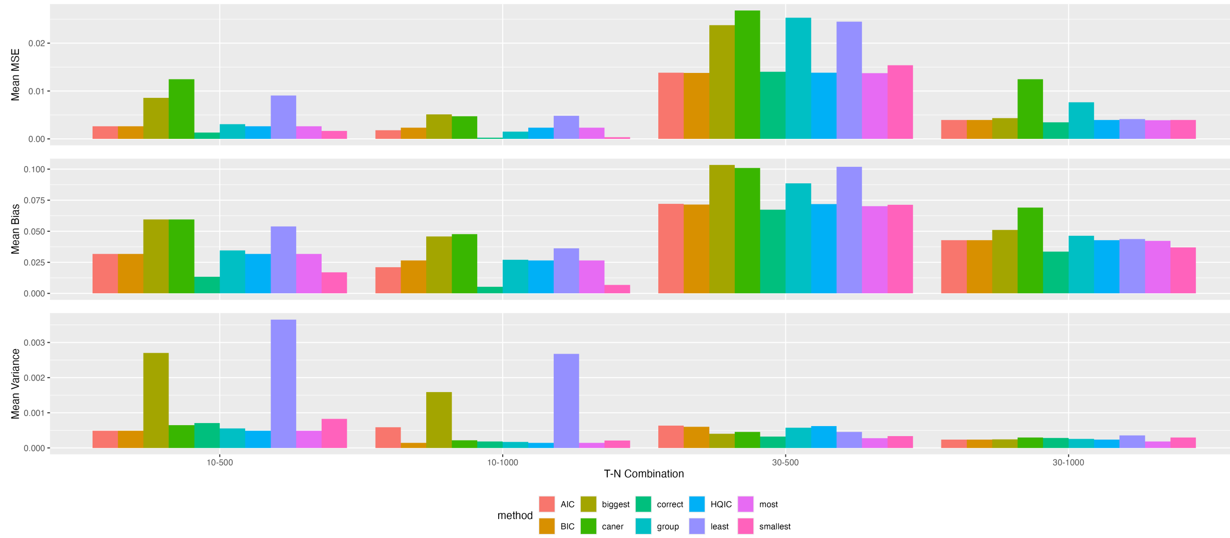


Figure 1: Performance metrics for stationary and homoskedastic models

5 Application: modeling the impact of mobility on covid spread

Covid-19 has been spreading in the U.S. since February 2020, infecting millions of Americans. To contain the situation, the government must control the transmission of the virus through public health policies until vaccines become available. From the onset of the pandemic, most countries adopted one of two primary strategies. The first strategy involved more relaxed measures aimed at achieving "herd immunity" by allowing the virus to spread through the population, with the hope of reaching collective immunity. In this approach, the policy measures typically focus on isolating infected individuals, promoting hygiene practices, and encouraging limited social distancing to mitigate the spread. The second strategy involved more stringent measures designed to cut off transmission routes and suppress the spread of the disease (Qiu et al., 2020). While most Western countries initially adopted the first approach, countries like China, South Korea, and Singapore implemented stricter policies from the outset. These countries adopted Nonpharmaceutical Interventions (NPI), which included restrictions on mobility, city lockdowns, mandatory quarantines, isolation of cases, and the closure of schools and non-essential businesses.

Although the latter approach is effective in controlling the spread, it is also controversial for both economic and political reasons. The restrictions on mobility have caused significant economic harm, including halted production, failing businesses, and rising unemployment.

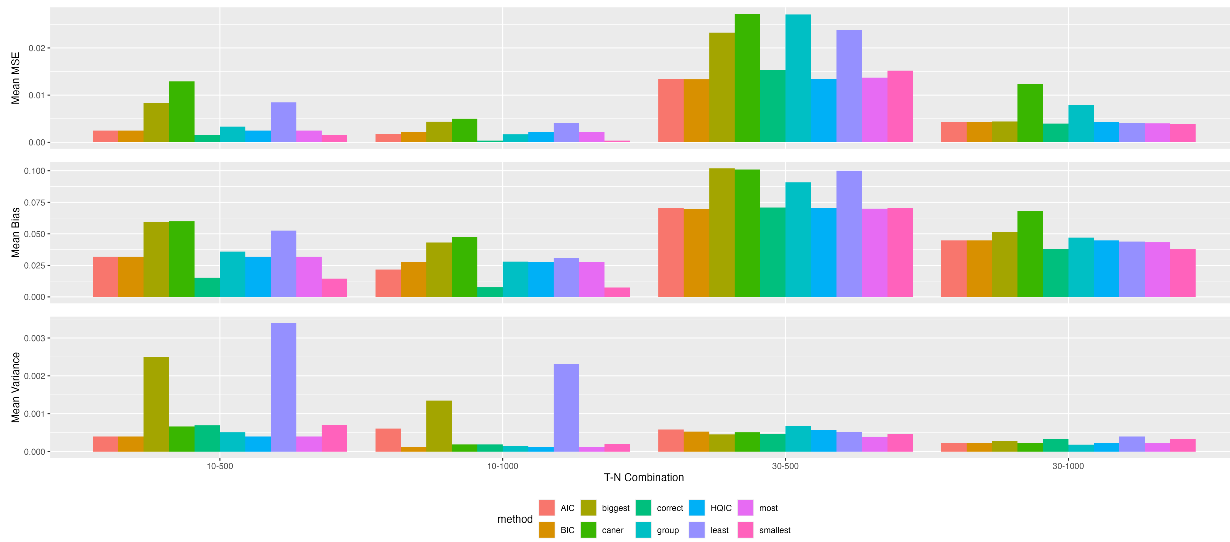


Figure 2: Performance metrics for non-stationary and homoskedastic models

Furthermore, the government's authority to restrict personal freedoms, such as mobility, is largely constrained by constitutional protections (Schwartz and Nathan, 2017). This trade-off has contributed to the slow policy response in the United States. Although the first confirmed case of Covid-19 was announced by the Centers for Disease Control on January 21, and the World Health Organization assessed the global risk of the coronavirus as "high" on January 27 (Presse, 2020), a national state of emergency was not declared until March 13, with the only mandatory national policy being international travel restrictions (Chowell and Mizumoto, 2020). Social distancing measures were not mandated at the federal or state levels until March 19, and even by April 8, five states had yet to implement social distancing rules, while three others had only partial rules in place (Sharkey, 2020).

Despite the challenges posed by mobility restrictions, most countries, including the U.S., eventually opted for this approach in the later stages of the pandemic for several reasons. First, during the early stages, contact tracing without widespread mobility restrictions remained viable, as testing and hospital capacity were sufficient to manage the relatively small number of infected individuals. However, in the later stages, when the virus had already spread across the country, reducing transmission fundamentally by limiting the pool of susceptible individuals became necessary to alleviate the burden on the healthcare system and keep the virus under control. Secondly, later medical research revealed that nearly half of all transmissions occur through pre-symptomatic and asymptomatic individuals (Ferretti et al., 2020). As shown in Figure 7 in the Appendix, which illustrates the estimated contributions to the basic reproduction rate (R0) from different transmission routes, this finding indicates

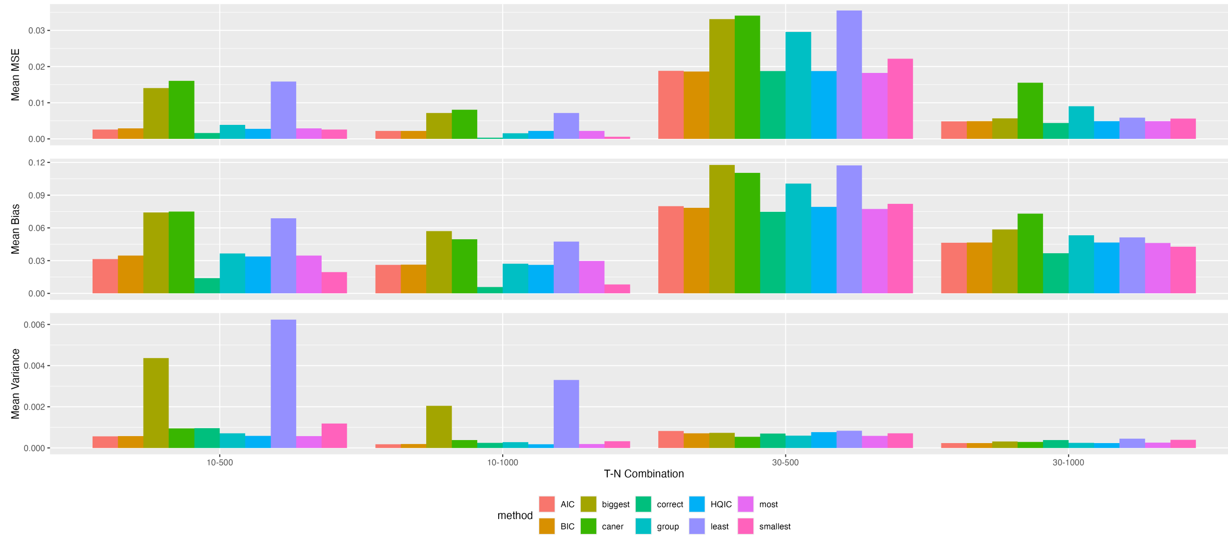


Figure 3: Performance metrics for stationary and heteroskedastic models

that tracing and quarantining only symptomatic individuals is insufficient to contain the disease. Therefore, broader societal measures restricting mobility are essential to controlling the spread.

Given that mobility restrictions are both costly and necessary for managing the current pandemic, it is crucial to evaluate and quantify their impact on the growth rate of new cases and death rates. Such an evaluation helps determine whether the benefits of these restrictions truly outweigh the economic, social, and political costs. To achieve this, it is essential to consider the methods available for assessing the effects of mobility on the spread of Covid-19. To my knowledge, the methods used in the literature on this subject can be divided into several categories: mechanistic, phenomenological, correlational, instrument variable-based, and event study-styled approaches. These methods provide different insights into the relationship between mobility restrictions and pandemic outcomes.

Mechanistic and phenomenological models are two key approaches to modeling the transmission process. Mechanistic models are based on the transmission process itself, incorporating parameters derived from epidemiological evidence, with others fitted to data. These models are particularly useful for counterfactual analysis and extrapolation beyond the estimation period. However, they rely heavily on assumptions that can be uncertain, especially in the context of an evolving pandemic, making them difficult to adapt to different stages of the outbreak. In contrast, phenomenological models estimate transmission through curve-fitting methods. While they lack the biological nuance of mechanistic models, their simplicity

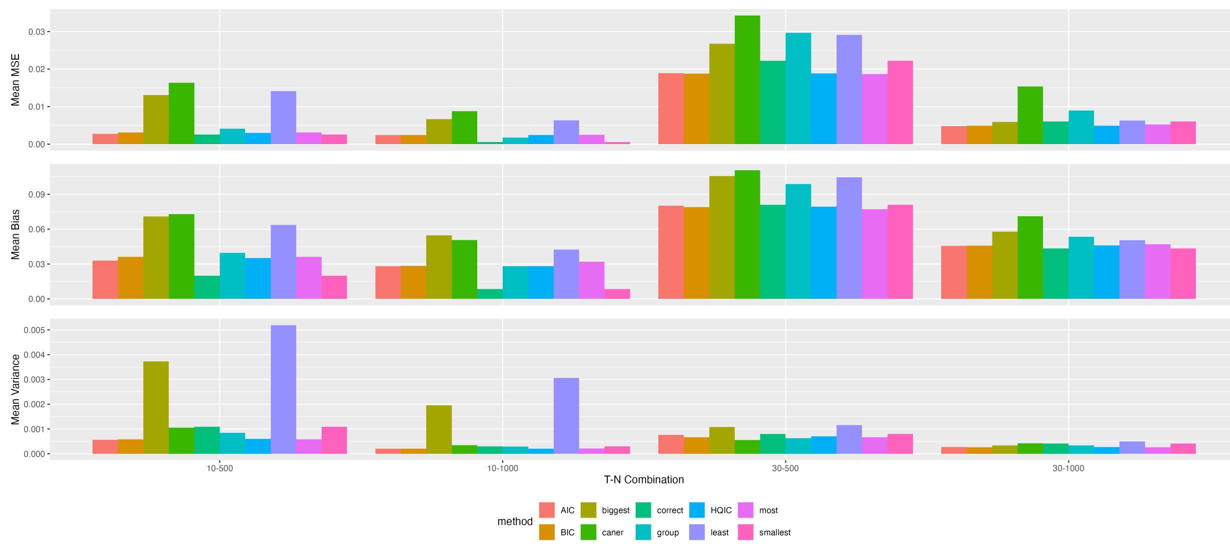


Figure 4: Performance metrics for non-stationary and heteroskedastic models

allows for easier comparison across different studies, which is often challenging with mechanistic models due to the variation in assumptions and processes they capture (Avery et al., 2020).

Examples of mechanistic models include those by Flaxman et al. (2020), Unwin et al. (2020), Hsiang et al. (2020), and Chen and Qiu (2020). Flaxman et al. (2020) studied the effects of major interventions across 11 European countries using a Bayesian model that simulates infection cycles, where different NPIs affect the time-varying reproduction number. Their results indicate that combined interventions lead to substantial reductions in the reproduction number. Similarly, Unwin et al. (2020) applied a similar methodology to U.S. states and found that increased mobility leads to a resurgence of transmission. Other studies, such as Hsiang et al. (2020), explored the effects of NPIs in various localities using a reduced-form econometric model based on the susceptible-infected-recovered (SIR) framework. However, their model's limited applicability in later stages of the pandemic and the short time span of their analysis present challenges for generalization.

Phenomenological models, such as those developed by Soucy et al., 2020, offer a different perspective. Their model used the Citymapper Mobility Index (CMI) to measure mobility in 41 cities worldwide, estimating a multilevel linear regression to study the association between mobility and case growth. While this method allows for broad comparisons, the lack of control for confounding variables like weather limits its ability to establish causal relationships. Another correlational study by Badr et al., 2020 used mobile phone data to correlate mobility with moving averages of case growth, but it faces similar limitations regarding causality.

Instrument variable-based approaches provide a different angle by using external factors to isolate the effect of mobility. For example, Kapoor et al. (2020) used rainfall before lockdown as an instrument for mobility, although their approach has been criticized for not accounting for incubation periods and the weak variation in mobility driven by rainfall. Glaeser et al. (2020) used employment by industry to measure mobility, leveraging differences in essential versus remote-work jobs, though this approach also suffers from potential omitted variable bias due to income disparities between areas.

Lastly, event study methods, such as Difference-in-Difference (DID) and Synthetic Control, have been employed to assess the impact of mobility restrictions on Covid-19 transmission. Research on U.S. data has yielded varied results depending on the data sources used. Abouk and Heydari (2020) found that strong measures like shelter-in-place orders (SIPO) significantly affected mobility and subsequently reduced case growth, while lenient policies had little impact. However, other studies, such as Gupta et al. (2020), argue that most reductions in mobility were voluntary rather than policy-driven, presenting a complex picture of how mobility restrictions influence transmission.

In terms of geographical focus, most studies have concentrated on China, where results consistently show a positive association between mobility and virus transmission. These studies measure mobility in different ways, including traffic in and out of Wuhan (Kraemer et al., 2020), social distancing measures (Tian et al., 2020), and national mobility flows (Chinazzi et al., 2020). Chinazzi et al. (2020) highlights the ripple effect of mobility changes in one location on global transmission dynamics, which points to the need for network modeling—though this approach is beyond the scope of this paper.

In this paper, I aim to evaluate the impact of mobility restrictions on the growth rate of Covid-19 cases by following a two-step approach. First, I estimate the relationship between mobility and Nonpharmaceutical Interventions (NPIs). Second, I analyze the effects of mobility on the growth rates of Covid-19 cases over several subsequent periods.

Step 1 focuses on estimating how different NPIs, such as "shelter-in-place" orders (SIPOs) and the closure of schools and non-essential businesses, affect mobility. Estimating these relationships using U.S. data is particularly advantageous due to the decentralized nature of policy decisions. The U.S. federal government delegated these decisions to individual states and counties, resulting in significant variations in decision-making processes, the types of measures implemented, and the timing of policy actions across the country (Adolph et al., 2020). This variation provides a rich dataset for examining how policy-induced reductions in mobility influence subsequent changes in the growth rate of Covid-19 cases. This analysis contributes to the literature focusing on comparing the effects of various NPIs on mobility (see Courtemanche et al., 2020 and Gupta et al., 2020).

Step 2, inspired by Wilson (2020), involves analyzing the effects of mobility on the growth rate of cases. Unlike Wilson (2020), however, this paper estimates the dynamic panel model using Generalized Method of Moments (GMM) and implements moment selection to address potential Nickell bias, which is not corrected for in the earlier work. By applying this method, I estimate how mobility levels during the pandemic influence case growth rates several periods ahead.

During the formulation of the model, several key considerations are taken into account. First, existing literature suggests heterogeneity in the effects of mobility on cases across different communities, based on factors such as income levels, education levels, racial compositions, population densities, and health expenditures (Wilson, 2020, Castex et al., 2020, Sa, 2020, Bonardi et al., 2020). To capture this variation, the model is run not only on the full dataset but also on subsets with different characteristics. The second key consideration involves selecting the appropriate dependent variable and lag structure. In this paper, I use two-week lagged effects of mobility on new Covid-19 cases, reflecting the time from infection to symptom onset (approximately 7 days) and the time from testing to confirmation of a case (another 7 days), as recommended by the medical literature (Lauer et al., 2020). Additionally, I control for county-specific time-invariant factors and nationwide time-varying issues in reporting by using county fixed effects and nationwide time fixed effects. However, potential biases may still arise from time-varying factors specific to certain counties, particularly if reporting quality varies at different stages of the pandemic. For example, reporting may be less accurate during the early stages or peak times of the pandemic, which could undermine the validity of the results.

In interpreting the findings, it is also important to be cautious. The effects of removing mobility restrictions may differ from the estimated effects of imposing them, as mobility could overshoot typical levels once restrictions are lifted due to pent-up demand for travel and social interactions.

5.1 Data

I obtained the daily U.S. county-level data on Covid-19 cases from the New York Times database. State-level data on daily testing are from the Covid Tracking Project, and these are obtained from the website tracktherecovery.org. The New York Times database records cumulative cases from the first reported coronavirus case in Washington State on Jan. 21, 2020, until the end of the study on Sep. 07, 2020. The data are collected from state and local governments and health departments. The cases include both those confirmed by laboratory RNA tests and probable ones based on clinical, epidemiologic, or serological testing, which mitigates the severe lack of RNA testing, especially at early stages. In Figure 8, the log cases for all counties over our sample period are presented. We can see that the number of cases started growing rapidly from mid-March for most counties, although some only started to record cases in May or June 2020. We can also observe large heterogeneity in the pattern of case development across counties during this period.

The Covid Tracking Project compiles testing data from each state for the period from Jan. 21, 2020, to Sep. 07, 2020, including both positive and negative results. The log total test data is shown in Figure 9. The number of tests began to take off around the beginning of March, and its growth across counties seems to be much more homogeneous, signaling concerted policy timings.

The mobility data we use is the Dallas Fed Mobility and Engagement Index (MEI). MEI data are calculated from SafeGraph data on the spatial behaviors of mobile devices, measuring at the county level information such as the fraction of devices leaving home per day, the fraction of devices away from home for 3-6 hours or longer than 6 hours, the fraction of devices traveling far or near from home, and the average daytime hours spent at home and away from home. These data are combined using principal component analysis to create the final indicator, MEI. This is seen as a measurement of mobility as well as engagement in economic activities. The available data range from Jan. 3 to Aug. 29, 2020. We present the time series of MEI for each county in Figure 10. We can observe that the mobility level decreased sharply from mid-March and only slowly recovered to its normal level around June 2020. This trend is shared uniformly by many counties, with varying degrees of magnitude in the drop in mobility.

The daily county-level weather data is constructed using the Global Historical Climatology Network Daily (GHCN-Daily) dataset. The data for the U.S. comes from the U.S. National Climate Data Center and include multiple measurements of weather features each day from over 15,000 weather stations located in the U.S. I constructed the weather variables for each county in a similar way as Wilson (2020). First, the contiguous U.S. is divided into 20-mile by 20-mile grids, and the weather features for each grid are calculated as the weighted average of readings from the weather stations within 50 miles, where the weights are the inverses of distances between the centroid of the grid and the weather stations. A demonstration of this approach is presented in Figure 5 and Figure 6. Note that not every station has data for each feature every day, so the daily data for these grids are taken as averages of available measurements on that day. Then the weather features for the counties are calculated as the averages of information from the grids within each county. The daily weather variables I collect span the period from Jan. 01 to Aug. 31, 2020, including measurements of precipitation, snowfall, and daily average temperature. The county-level time series for each weather variable is shown in Figures 12, 13, and 14.

Figure 5

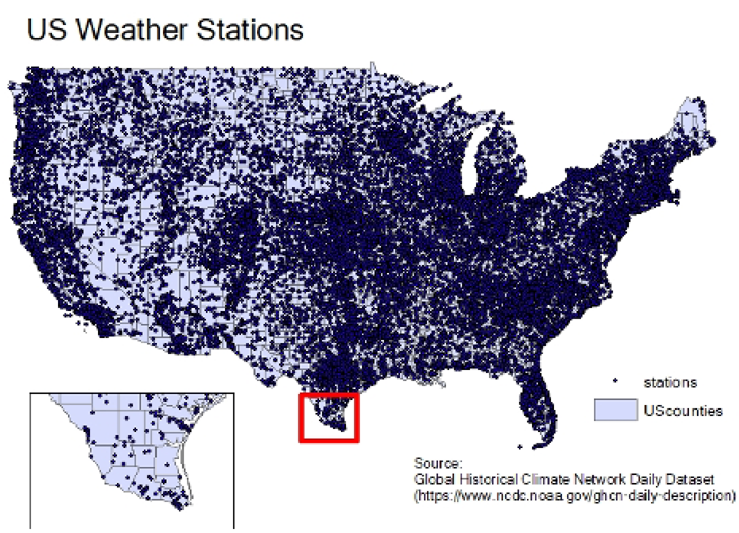
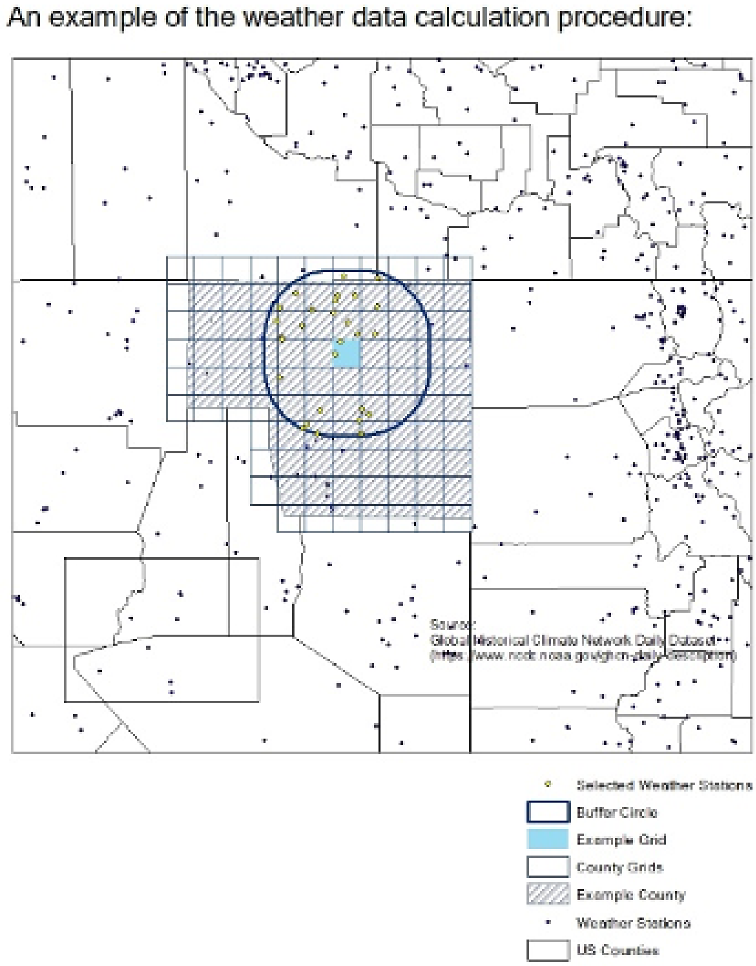


Figure 6



Data on the NPIs are from the Keystone Coronavirus City and County Non-Pharmaceutical Intervention Rollout Date Dataset. The data are hand-collected from government health websites and local news reports, and therefore, they are constrained in terms of the size and time period they cover. The dataset records 618 county-level NPIs and 53 state-level NPIs, with the starting dates of NPIs ranging from Mar. 08 to Jun. 08, 2020, and the ending dates ranging from Mar. 14 to Jun. 27, 2020. Although the number of localities this dataset covers is small, it is said to cover all U.S. states and counties with at least 100 confirmed cases as of April 06, 2020. The NPIs are grouped into 11 categories: closing of public venues, gatherings of size 0-10, 11-25, 26-100, 101-500, lockdown, non-essential service closure, religious gatherings banned, school closure, shelter-in-place, and social distancing. Figure 11 presents the occurrence of various NPIs for each county during the sample period. Each plot represents an NPI, and the rows are counties and columns are dates. Each row is colored red for the period when the NPI is active. We can observe from the plot that various NPIs were introduced around mid-March. There are noticeable variations in the adoption time across different counties for stricter NPIs, including shelter-in-place orders, non-essential services closure, and banning gatherings of size 11 to 25. For other NPIs like school closures and banning gatherings of size 101 to 500, the action times were much more uniform and in effect for a longer period.

For the control variables, I collect the following, similar to suggestions in previous literature. To begin with, I have data on the average number of residents per day for each county in September 2020 as an indicator of the health capacity of the county, from the Federal Centers for Medicare & Medicaid Services. It can be seen from Figure 15 that this variable follows a skewed distribution, with most counties having a similar number of residents and some counties having much more. To evaluate the heterogeneous effects for counties with different demographic characteristics, I also collect county-level data from the U.S. Department of Agriculture Economic Research Service on the percentage of the adult population with an education level of at most high school/at least a bachelor's degree during 2014-2018, the population level as of 2019, net domestic and international migration rates from 2018 to 2019, the percentage of all people and of people aged 0-17 in poverty in 2018, the median household income in 2018, and the unemployment rate in 2019. Detailed graphs on these demographic characteristics can be seen in Figure 15. In addition, due to the large overlap area of our NPIs, as can be seen from Figure 11, there are high covariance between the NPI policy variables. In order to avoid the multicollinearity problem and improve interpretation, we use principal componnet analysis to extract the first 4 principal component from our 11 NPI variables. These explain over 80% of all the variations in our NPI data. We present in Figure 16 the loadings of each NPI variable on the principal components. It's clear that the first component measures the global effect of all NPIs (with the opposite sign). The second component loads mainly on lockdown and banning gathering of size 11 to 25 (with the correct sign). The third component represents (with the opposite sign) the banning of gatherings of size 101 to 500. Lastly, the fourth component represents mainly the banning of gatherings of size 26 to 100 (with a positive sign).

5.2 Model

The first stage is to study the relationship between NPIs and mobility. I first aggregate the daily data into a weekly panel dataset for both the first and second stage models because the

Covid-19 data suffers from lagged reporting and occasional corrections, which usually occur within the next few days. These corrections often happen within a week, so the weekly data may be more reliable than daily data.

$$m_{it} = \beta_1 g_{it} + \beta_2 g_{it-1} + w_{it}^T \beta_3 + PC_{it}^T \theta + \alpha_i + \alpha_t + \varepsilon_{it}, \tag{61}$$

where m_{it} is the MEI index, w_{it} is the vector of weather variables (precipitation, snowfall, and average temperature), and x_{it} is the testing growth. PC_{it} represents the principal components of the NPIs adopted. α_i and α_t control for county and time fixed effects that may affect both mobility and the explanatory variables, such as county-specific demographic characteristics or national-level policy guidance. This is important, as there tend to be national trends in weather variables, and the time fixed effects can capture this.

The second stage of the model estimates the weekly dynamic panel data model:

$$g_{it} = \alpha_1 g_{it-1} + \alpha_2 g_{it-2} + \phi_1 m_{it-1} + \phi_2 m_{it-2} + w_{it-1}^T \delta_1 + w_{it-2}^T \delta_2 + \tau x_{it} + \alpha_i + \alpha_t + \varepsilon_{it}, \tag{62}$$

where g_{it} is the weekly growth rate of new cases, m_{it} is the weekly averaged mobility measurement, w_{it} is the vector of weekly-averaged weather variables, and x_{it} is the testing growth this period. Two-period lags of mobility and weather are included because they affect disease transmission and the probability of infection. The average period from infection to confirmed cases is usually two weeks, but many individuals show symptoms and get tested more quickly. Mobility is assumed to be predetermined but may not be strictly exogenous, as people's willingness to go out might be influenced by current and previous case growth. Weather variables are taken as exogenous, as in previous literature. Testing growth is included as it affects how many infected cases are recorded, and this variable is considered exogenous after controlling for county-level fixed effects, which capture differences in health capacity.

County and time fixed effects are added to the model to account for unobserved fixed characteristics, as well as county- or time-specific measurement errors, which are a significant issue due to the current poor quality of data. To estimate the model using the Sparse Group Lasso approach, I first demean the variables across counties to remove the time-constant effects, then set up the sure and unsure moment conditions depending on the type of variables, and estimate the model using the sparse group lasso approach proposed in Section 3.

To get standard errors on our estimated coefficients, we apply bootstrapping on the counties. As discussed in Kapetanios (2008), when there is temporal dependence in the data, bootstrapping with iid cross-sectional data gives a more accurate distribution approximation, compared with temporal bootstrapping. This is because the former keeps the time-series structure intact. We follow this suggestion and bootstrap samples across counties. However, counties within the same state may still exhibit weak dependence. Therefore, we sample on the state level to keep the correlation structure inside states intact. To make the resampled data comparable with the original data in terms of the state-county structure, we only resample states within the cluster of states with the same number of counties. This is the idea proposed in Sherman and Cessie (1997). For each estimate, we resampled 1000 times. We use the mean estimate as our final estimate and the standard deviation of the estimate as our estimated standard error.

5.3 Results and implications

First-stage results For the estimation of the first-stage model, 10 regressions are estimated, including the baseline model and the results for 9 subsamples. These results are reported in Tables 4, 3, and 2, and all the regressions are estimated using panel data models with county fixed effects (after we first demean across counties to eliminate the time fixed effects). We used Driscoll-Kraay HC3 clustered standard errors Hoechle (2007) to account for the cross-sectional dependence of counties and temporal dependence within each county.

In the baseline model, both current and lagged case growth, as well as test growth, have a significant negative effect on mobility. This is intuitive as people voluntarily reduce mobility to avoid getting infected. Higher temperature encourages mobility, while higher precipitation discourages it. In terms of the NPIs, the average effect of all NPI measures is negative and strongly significant. Strict measures like lockdowns and banning gatherings of size 11 to 25 have a significant negative effect on mobility, but the magnitude is only one-third of the average effect. This is perhaps because when these strict measures are taken, the Covid situation is already severe, and voluntary decreases in mobility dominate, resulting in a smaller additional effect of the NPIs. Notice that PC3, which represents banning gatherings of size 101 to 500, has a significant and large negative effect. This could be due to the fact that this relatively mild NPI started early in the pandemic. This range of group size corresponds to large gatherings like concerts or conferences, both of which involve a lot of long-distance travel. This policy tool is then very effective in reducing mobility. Somewhat surprising is the effect of PC4, which represents the banning of gatherings of size 26 to 100. This effect is significant but positive. This shows there may be some substitutions from gatherings of this size to smaller gatherings, which ultimately increase mobility.

The results from different subgroups are mostly consistent with the above observations despite a few exceptions. Interestingly, banning large group gatherings of size 101 to 500 has no effect in poorer areas (seen from Table 2) as it's less likely to host large events. They also respond to PC4 with a larger positive effect, which could point to a lack of regulation in these areas and large substitution effects of the NPI. Strict measures like lockdowns are more effective for counties with more domestic migration (seen from Table 3), as domestic migrants tend to be more mobile and need more travel to maintain social ties. Strict measures greatly reduce the possibilities. From Table 4, we also notice that lockdowns are more effective in reducing mobility for less-educated areas than for more-educated areas, possibly due to a higher proportion of people engaging in manual or service-related work. Banning gatherings of size 26 to 100 is again positive and significant for less-educated areas, signaling a substitution effect similar to what was observed in poorer areas.

From these results, we learn that most NPIs are effective in reducing mobility, but the right NPI choice for each area needs to take into account local demographics. What works for one location may not be effective in another and may even have unintended policy effects as we discussed above.

Second-stage results The second-stage results are reported in Table 5. We formed our standard error using bootstrapped samples, and the test statistics are calculated as the ratio between the mean estimates and the standard deviation of the bootstrapped estimates. We can see that the Covid case growth rate is significantly affected by the previous two lags. Among the rest of the insignificant variables, temperature and mobility from two periods ago have the largest test statistics. This supports our belief that it takes around two weeks for the effect of infections to show up as new cases. Mobility has a positive effect on cases, whereas higher temperature reduces cases. The results are more interesting when we look at different subsamples in our dataset (see the results in Table 6). Past case growth shows a consistently significant effect on current case growth across most groups, particularly in areas of large population. In terms of mobility, only mobility from two periods ago has significantly positive effects on case growth for some subsamples. The effect is particularly noticeable for less-educated areas, areas with a high proportion of international migration, and highly populated areas. We also see some evidence of precipitation increasing the case growth rate (also with two-period lags), supporting the argument that humidity makes the Covid virus last longer. Temperature has a large negative and significant effect on case growth, which is consistent with common beliefs from medical research. These results show that controlling mobility is effective (at least in some demographics) for controlling Covid case growth, even after controlling for important influencing factors like temperature and precipitation.

6 Conclusion

In this paper, I focus on estimating dynamic panel models using the GMM method. I begin by discussing the available moment conditions under various combinations of assumptions about the variables and error structures. The specific grouping structure of these moment conditions motivates the use of the Sparse Group Lasso approach. A simulation study was conducted to compare the performance of this method against benchmark methods. AL methods generally perform the best under the simulation settings. Sparse Group Lasso yields results that are close to the AL methods in most cases, except in high-dimensional settings where shrinkage bias worsens its performance. Despite the superior performance of AL methods, they are computationally expensive as the number of candidate models and moment conditions increases. Compared with the adaptive elastic net method from Caner et al. (2018), I find that the Sparse Group Lasso approach exhibits better properties in terms of bias, standard errors, and RMSE of the estimators.

In the empirical section, we estimate the effects of different NPIs on reducing mobility, demonstrating that this effect varies depending on local demographics. Mobility is also shown to affect Covid case growth in certain subsamples. Areas with these features should be cautious in regulating mobility, as the persistence in Covid case growth implies that once there is a surge, it may continue for multiple periods. The results of the empirical analysis can help inform better policymaking to contain the Covid pandemic.

Table 2: First-stage Regression results for MEI (Poverty, Income, and Unemployment)

	Dependent variable:			
			MEI	
	Baseline	Poverty (more)	Median Household Income (less)	Unemployment (more)
	(1)	(2)	(3)	(4)
casegrowth	−0.020***	−0.019***	−0.018***	−0.018***
	(0.004)	(0.005)	(0.005)	(0.006)
lag(casegrowth)	−0.040***	−0.034***	−0.032***	−0.033***
	(0.004)	(0.006)	(0.005)	(0.006)
Precipitation	−0.024***	−0.024***	−0.025***	−0.026***
	(0.002)	(0.002)	(0.002)	(0.002)
Snow	0.105	−0.0001	−0.058	0.012
	(0.111)	(0.173)	(0.118)	(0.146)
Temperature	0.077***	0.089***	0.071***	0.078***
	(0.005)	(0.007)	(0.007)	(0.007)
testgrowth	−0.047***	−0.070***	−0.074***	−0.055**
	(0.014)	(0.017)	(0.017)	(0.024)
PC1	0.635***	0.497***	0.484***	0.565***
	(0.073)	(0.092)	(0.091)	(0.104)
PC2	−0.193**	−0.228**	−0.341**	−0.190
	(0.083)	(0.116)	(0.145)	(0.116)
PC3	0.586***	0.245	0.204	0.566**
	(0.156)	(0.210)	(0.219)	(0.222)
PC4	0.594***	0.941***	1.238***	0.836**
	(0.221)	(0.313)	(0.359)	(0.375)
Observations	14,525	7,199	7,253	6,659
R^2	0.176	0.196	0.166	0.178
Adjusted R^2	0.127	0.148	0.116	0.129
F Statistic	292.051*** (df = 10; 13708)	165.939*** (df = 10; 6789)	136.339*** (df = 10; 6840)	136.166*** (df = 10; 6279)

Note: *p<0.1; **p<0.05; ***p<0.01

Note: The standard errors are calculated using Driscoll-Kraay HC3 clustered standard errors and each column includes ETF-specific fixed effects.

Table 3: First-stage Regression results for MEI (Population, Internation and Domestic Migration)

	Dependent variable:MEI			
	Baseline	Population (more)	International Migration (more)	Domestic Migration (more)
	(1)	(2)	(3)	(4)
casegrowth	−0.02	−0.02**	−0.01	−0.03***
	(0.31)	(0.01)	(0.01)	(0.01)
lag(casegrowth)	−0.04	−0.07***	−0.04***	−0.04***
	(0.24)	(0.01)	(0.01)	(0.01)
Precipitation	−0.02	−0.02***	−0.02***	−0.03***
	(0.00)	(0.00)	(0.00)	(0.00)
Snow	0.11***	0.65***	−0.01	−0.04
	(0.04)	(0.25)	(0.17)	(0.15)
Temperature	0.08***	0.10***	0.09***	0.06***
	(0.00)	(0.01)	(0.01)	(0.01)
testgrowth	−0.05	0.01	−0.01	−0.06***
	(0.65)	(0.02)	(0.02)	(0.02)
PC1	0.64***	0.60***	0.77***	0.51***
		(0.09)	(0.10)	(0.10)
PC2	−0.19***	−0.18**	−0.14	−0.45***
		(0.09)	(0.11)	(0.13)
PC3	0.59***	0.67***	0.76***	0.15
		(0.17)	(0.19)	(0.20)
PC4	0.59	0.71***	0.82***	0.01
		(0.22)	(0.25)	(0.28)
Observations	14,525	7,253	7,253	7,253
R^2	0.18	0.25	0.21	0.14
Adjusted R^2	0.13	0.20	0.16	0.09
F Statistic	292.05*** (df = 10; 13,708)	223.91*** (df = 10; 6,840)	183.56*** (df = 10; 6,840)	115.20*** (df = 10; 6,840)

*p<0.1; **p<0.05; ***p<0.01

Note: The standard errors are calculated using Driscoll-Kraay HC3 clustered standard errors and each column includes ETF-specific fixed effects.

Table 4: First-stage Regression results for MEI (Residents and Education)

	Dependent variable: MEI			
	Baseline	Residents (fewer)	High School (more)	Bachelor (more)
	(1)	(2)	(3)	(4)
casegrowth	−0.02	−0.02***	−0.02***	−0.02***
	(0.31)	(0.01)	(0.01)	(0.01)
lag(casegrowth)	−0.04	−0.03***	−0.04***	−0.05***
	(0.24)	(0.01)	(0.01)	(0.01)
Precipitation	−0.02	−0.03***	−0.02***	−0.02***
	(0.00)	(0.00)	(0.00)	(0.00)
Snow	0.11***	−0.25	−0.26	0.30*
	(0.04)	(0.18)	(0.27)	(0.16)
Temperature	0.08***	0.05***	0.08***	0.09***
	(0.00)	(0.01)	(0.01)	(0.01)
testgrowth	−0.05	−0.08***	−0.07***	−0.01
	(0.65)	(0.02)	(0.02)	(0.02)
PC1	0.64***	0.66***	0.71***	0.70***
		(0.12)	(0.11)	(0.10)
PC2	−0.19***	−0.13	−0.39*	−0.12
		(0.25)	(0.21)	(0.10)
PC3	0.59***	0.09	0.49	0.69***
		(0.40)	(0.32)	(0.18)
PC4	0.59	−0.61	1.39***	0.45*
		(0.56)	(0.41)	(0.24)
Observations	14,525	7,253	7,235	7,235
R^2	0.18	0.13	0.20	0.21
Adjusted R^2	0.13	0.07	0.15	0.16
F Statistic	292.05*** (df = 10; 13,708)	98.27*** (df = 10; 6,840)	167.32*** (df = 10; 6,823)	179.99*** (df = 10; 6,823)

*p<0.1; **p<0.05; ***p<0.01

Note: The standard errors are calculated using Driscoll-Kraay HC3 clustered standard errors and each column includes ETF-specific fixed effects.

Table 5: Second stage results (Full sample)

Variable	param_mean	param_sd	param_tstat
lag(casegrowth)	0.258 76	0.088 09	2.9374 ***
lag(casegrowth,2)	0.146 37	0.061 09	2.3962 **
lag(mobility)	0.010 33	0.031 98	0.322 89
lag(mobility,2)	0.070 78	0.067 34	1.051
lag(precipitation)	0.013 90	0.050 73	0.274 09
lag(precipitation,2)	0.002 58	0.003 56	0.726 77
lag(snow)	0.000 44	0.001 96	0.226 89
lag(snow,2)	0.159 09	0.361 51	0.440 07
lag(temperature)	−0.128 75	0.296 91	−0.433 64
lag(temperature,2)	−0.013 45	0.008 33	−1.6147
testing	−0.004 13	0.007 75	−0.533 08

Note: *p<0.1, **p<0.05, ***p<0.01.

7 Appendix

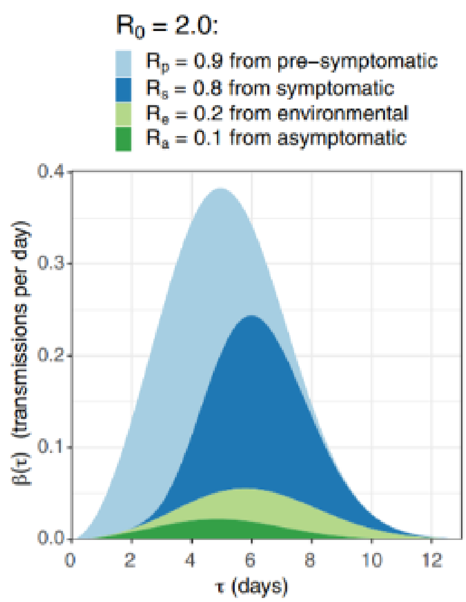


Figure 7: R_0

References

Abouk, R. and Heydari, B. (2020). The Immediate Effect of COVID-19 Policies on Social Distancing Behavior in the United States. Retrieved from medRxiv.

Table 6: Second stage results (Heterogeneous effects)

Variable	**Bachelor**	**Domestic**	**Income**	**High-school**	**International**	**Population**	**Poverty**	**Resident**	**Unemployed**
lag(casegrowth)	2.5253 ***	1.3142 **	0.950 95	0.755 55	1.918 **	3.0015 ***	1.0986	0.409 76	−0.180 42
lag(casegrowth,2)	1.5332 ***	1.6786 ***	1.0323 **	1.1498 **	2.3356 ***	1.4633 **	1.0251	0.6513	0.504 23
lag(mobility)	0.192 71	−0.007 51	0.186 08	0.479 49	0.500 22	0.774 68	0.6026	−0.3176	0.264 37
lag(mobility,2)	0.819 12**	0.701 47	0.853 41*	1.2199 **	1.721 **	2.020 ***	0.855 48	0.297 99	0.444 27
lag(precipitation)	−0.386 34	0.253 41	0.182 15	0.237 65	0.349 76	−0.1059	0.070 29	0.502 72	0.278 88
lag(precipitation,2)	0.826 87*	1.0743 **	0.415 66	0.3412	0.550 99	0.682 74**	0.436 66	0.639 12	0.165 61
lag(snow)	0.486 61	0.533 79	−0.2425	0.115 65	0.198 15	0.841 65	−0.1770	−0.469	−0.176 02
lag(snow,2)	0.462 79	−0.1256	0.386 67	0.450 97	0.468 94	−0.1001	0.404 76	−0.1463	0.399 37
lag(temperature)	−0.168 33	0.239 87	0.420 33	0.048 28	−0.2459	0.382 05	0.328 95	−0.7918	0.209 38
lag(temperature,2)	−1.2741 ***	−1.0503 ***	−1.0745 ***	−0.736 77**	−1.7671 ***	−1.0276 **	−1.201	−1.355	−1.1747
testing	−0.617 56	−0.334 65	−0.260 86	−0.5737	−1.0114	−0.619 46	−0.197 03	0.278 39	0.4452

Note: *p<0.1, **p<0.05, ***p<0.01. The values in the above table are test statistics that are formed dividing mean parameter estimates by their standard deviations.

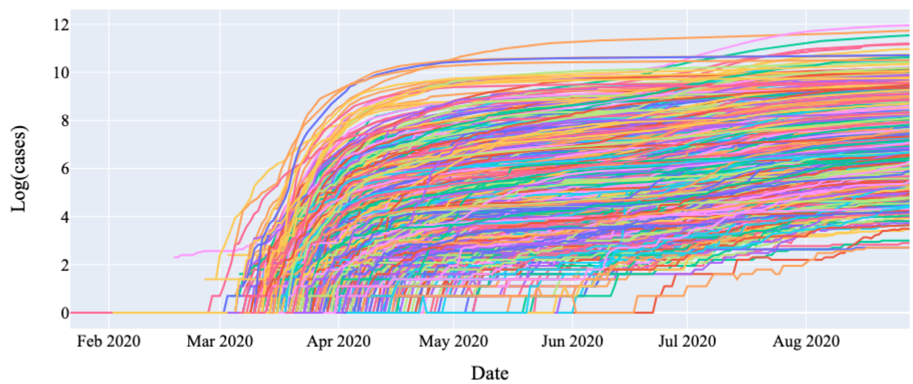


Figure 8: Log(cases+1) over time and across counties

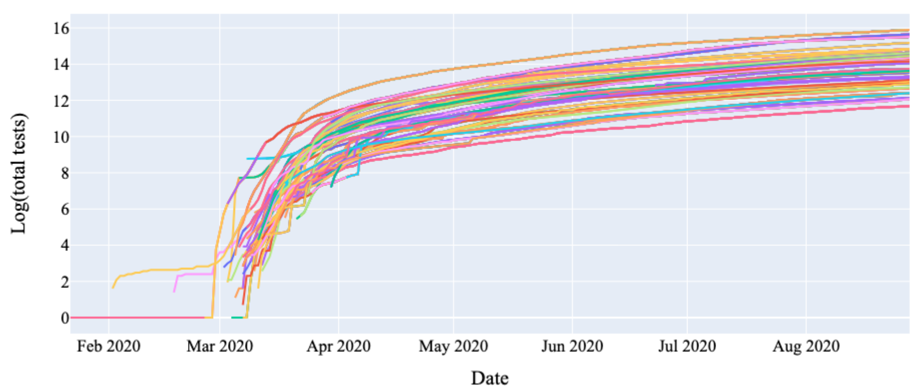


Figure 9: Log(total tests+1) over time and across counties

Adolph, C., Amano, K., Bang-Jensen, B., Fullman, N., and Wilkerson, J. (2020). Pandemic politics: Timing state-level social distancing responses to covid-19. Journal of Health Politics, Policy and Law, 8802.

Ahn, S. C. and Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. Journal of Econometrics, pp, 68:5–27.

Andrews, D. W. (1999). Consistent moment selection procedures for generalized method of moments. Econometrica, pp, 67(3):543–564.

Andrews, D. W. and Lu, B. (2001). Consistent model and moment selection procedures for

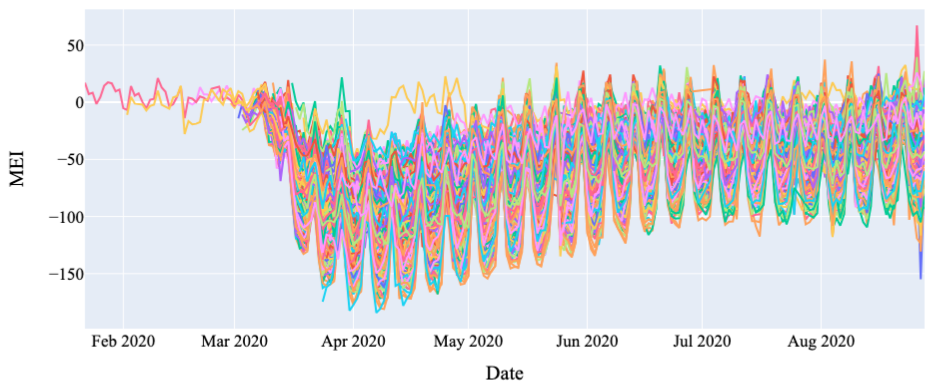


Figure 10: Time series of mobility index MEI across counties

gmm estimation with application to dynamic panel data models. Journal of Econometrics, pp, 101:123–164.

Arellano, M. and Bond, S. n. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. In The Review of Economic Studies, pages 277–297. pp. Vol. 58, No. 2 (Apr Published.

Avery, C., Bossert, W., Clark, A., Ellison, G., and Ellison, S. F. (2020). Policy implication of models of the spread of coronavirus: Perspectives and opportunities for economists. Working Paper 27007, NATIONAL BUREAU OF ECONOMIC RESEARCH, p.

Badr, H., Du, H., Marshall, M., Dong, E., Squire, M., and Gardner, L. M. (2020). Social Distancing is Effective at Mitigating COVID-19 Transmission in the United States. Retrieved from medRxiv.

Belloni, A. P., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica, pp, 80(6):2369–2429.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, pp, 87:115–143.

Bonardi, J.-P., Gallea, Q., Kalanoski, D., and Lailve, R. (2020). Fast and local: How lockdown policies affect the spread and severity of covid-19. Covid Economics, pp, 23:325–351.

Bond, S. (2002). Dynamic panel data models: A guide to micro data methods and practice. cemmap. Technical report, working paper, p. No. CWP09/02.

Caner, M., Han, X., and Lee, Y. (2018). Adaptive elastic net gmm estimation with many invalid moment conditions: Simultaneous model and moment selection. Journal of Business Economic Statistics, pp, 36:1.

Castex, G., Dechter, E., and Lorca, M. (2020). Covid-19: Cross-country heterogeneity in effectiveness of non-pharmaceutical interventions. Covid Economics, pp, 14:175–199.

Chen, X. and Qiu, Z. (2020). Scenario analysis of non-pharmaceutical interventions on global covid-19 transmissions. Covid Economics, pp, 7:46–67.

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., and Litvinova, M. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. Science, pp, 368:395–400.

Chowell, G. and Mizumoto, K. (2020). The covid-19 pandemic in the usa: what might we expect? The Lancet, 395(10230):1093–1094.

Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., and Yelowitz, A. (2020). Strong social distancing measures in the united states reduced the covid-19 growth rate. HEALTH AFFAIRS, pp, 39.

Ferretti, L., Wymant*, C., Kendall, M., Zhao, L., Nurtay, A., and Abeler-D"orner, L. (2020). . . Fraser, C. Quantifying SARS-CoV-, 368:6491.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J., Mellan, T. A., and Coupland, H. (2020). . . Bhatt, S. Estimating the effects of non-pharmaceutical interventions on COVID-, 19.

Forbes, K. J. (2000). A reassessment of the relationship between inequality and growth. American Economic Review, pp, 90:869–887.

Fujiwara, Y., Ida, Y., Shiokawa, H., and Iwamura, S. (2016). Fast lasso algorithm via selective coordinate descent. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, pages 1561–1567.

Glaeser, E. L., Gorback, C. S., and Redding, S. J. (2020). How much does covid-19 increase with mobility? evidence from new york and four other u. Working Paper 27519, S. CITIES. NATIONAL BUREAU OF ECONOMIC RESEARCH, p.

Gørgens, T., Han, C., and Xue, S. (2019). Moment restrictions and identification in linear dynamic panel data models. Annals of Economics and Statistics, pp, 134:149–176.

Gupta, S., Nguyen, T. D., Rojas, F. L., Raman, S., Lee, B., and Bento, A. (2020). . . Wing, C. TRACKING PUBLIC AND PRIVATE RESPONSES TO THE COVID-, 19.

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. The stata journal, 7(3):281–312.

Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., and Bolliger, I. (2020). The effect of large-scale anti-contagion policies on the coronavirus (covid-19) pandemic. Nature, pp, 584:262–267.

Judson, R. A. and Owen, A. L. (1999). Estimating dynamic panel data models: a guide for macroeconomics. Economics Letters, pp, 65:9–15.

Kang, H., Zhang, A., Cai, T. T., and Small, D. S. (2020). Instrumental variables estimation with some invalid instruments and its application to mendalian randomization. Journal of the American Statistical Association, pp, 111:513.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. The Econometrics Journal, 11(2):377–395.

Kapoor, R., Rho, H. A., Sangha, K., Sharma, B., Shenoy, A., and Xu, G. (2020). God is in the Rain: The Impact of Rainfall-Induced Early Social Distancing on COVID-19 Outbreaks. Retrieved from SSRN.

Kraemer, M. U., Yang, C.-H., Gutierrez, B., Wu, C.-H., and Klein, B. (2020). The effect of human mobility and control measures on the covid-19 epidemic in china. Science, pp, 368:493–497.

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., and Zheng, Q. (2020). The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application. Ann Intern Med, pp, pages M20–0504.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of multivariate analysis, 88(2):365–411.

Levine, R., Loayza, N., and Beck, T. (2000). Financial intermediation and growth: causality and causes. Journal of Monetary Economics, pp, 46:31–77.

Nickell, S. (1981). Biases in dynamic models with fixed effects. Econometrica, pp, 49:1417–1426.

Noah, S., Friedman, J., Hastie, T., and Tibshirani, R. (2013). A sparse group lasso. Journal of Computational and Graphical Statistics, pp, 22:231–245.

Presse, A. F. (2020). WHO says global risk of China virus is 'high'. Retrieved from TheStar.

Qiu, Y., Chen, X., and Shi, W. (2020). Impacts of social and economic factors on the transmission of coronavirus disease (covid-19) in china. Journal of Population Economics , pp, 33:1127–1172.

Roodman, D. (2009). Practitioners. Corner. A Note on the Theme of too many Instruments'. Oxford Bulletin of Economics and Statistics, 71:135–58.

Sa, F. (2020). Socioeconomic determinants of covid-19 infections and mortality: Evidence from england and wales. Covid Economics, pp, 22:47–58.

Schwartz, B. and Nathan, N. C. (2017). Choice, freedom, and well-being: considerations for public policy. Behavioural Public Policy, pp, 1(1):106–121.

Sharkey, P. (2020). The us has a collective action problem that's larger than the coronavirus crisis. Vox. https://www. vox. com/2020/4/10/21216216/coronavirus-social-distancing-texas-unacast-climate-change.

Sherman, M. and Cessie, S. l. (1997). A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models. Communications in Statistics-Simulation and Computation, 26(3):901–925.

Soucy, J.-P. R., Sturrock, S. L., Berry, I., Daneman, N., MacFadden, D. R., and Brown, K. A. (2020). Estimating the effect of physical distancing on the COVID-19 pandemic using an urban mobility index. Retrieved from medRxiv.

Tian, H., Liu, Y., Li, Y., Wu, C.-H., Chen, B., and Kraemer, M. U. (2020). An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in china. Science, pp, 368:638–642.

Unwin, H. J., Mishra, S., Bradley, V. C., Gandy, A., Mellan, T. A., and Coupland, H. (2020). State-level tracking of covid-19 in the united states. Imperial College COVID-, 19.

Wilson, D. J. (2020). Weather, social distancing, and the spread of covid-19. Federal Reserve Bank of San Francisco Working Paper.

Windmeijer, F., Farbmacher, H., Davies, N., and Smith, G. D. (2020). On the use of the lasso for instrumental variables estimation with some invalid instruments. Journal of the American Statistical Association, pp, 114:527.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society, p. Series B, 68:49.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, pp, 67:301–320.

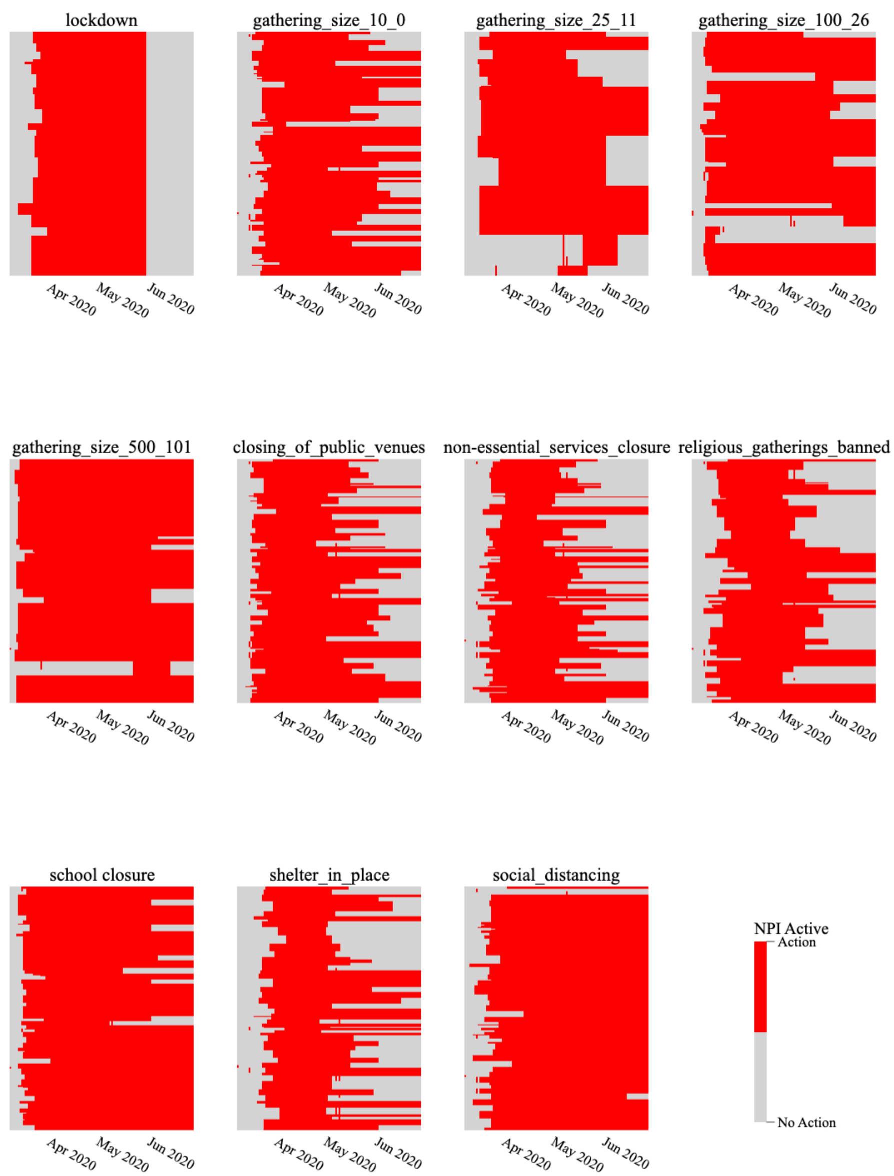


Figure 11: Occurrence of various NPIs across counties (rows) and over time (columns)

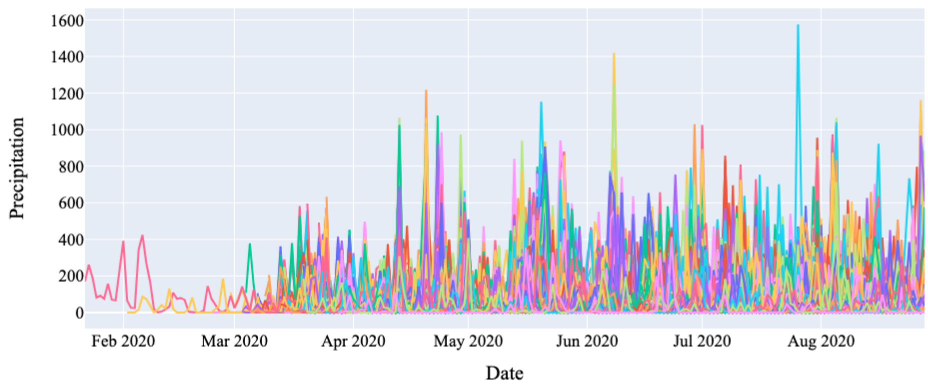


Figure 12: Time series of precipitation for all sample counties (tenths of a millimeter)

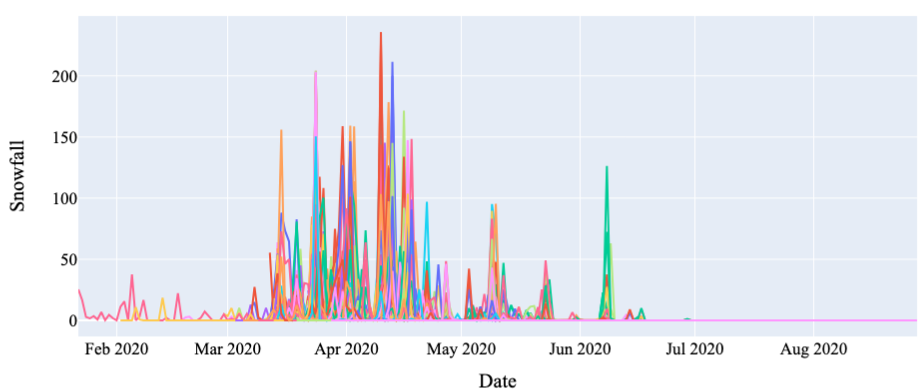


Figure 13: Time series of snowfall for all sample counties (millimeters of snow depth)

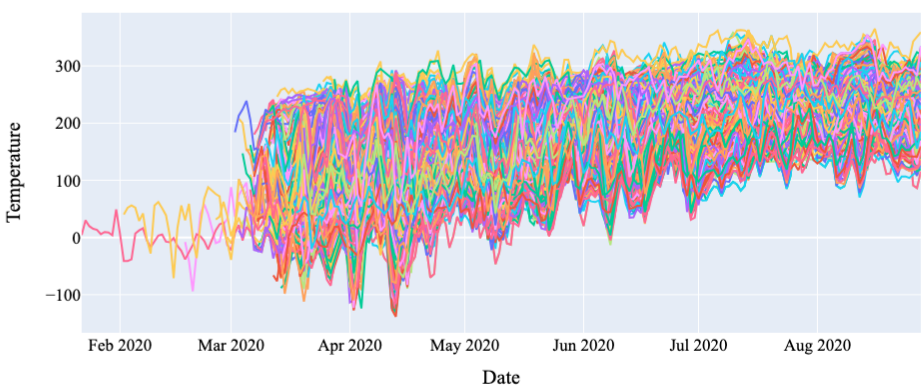


Figure 14: Time series of temperature for all sample counties (tenths of degrees Celsius)

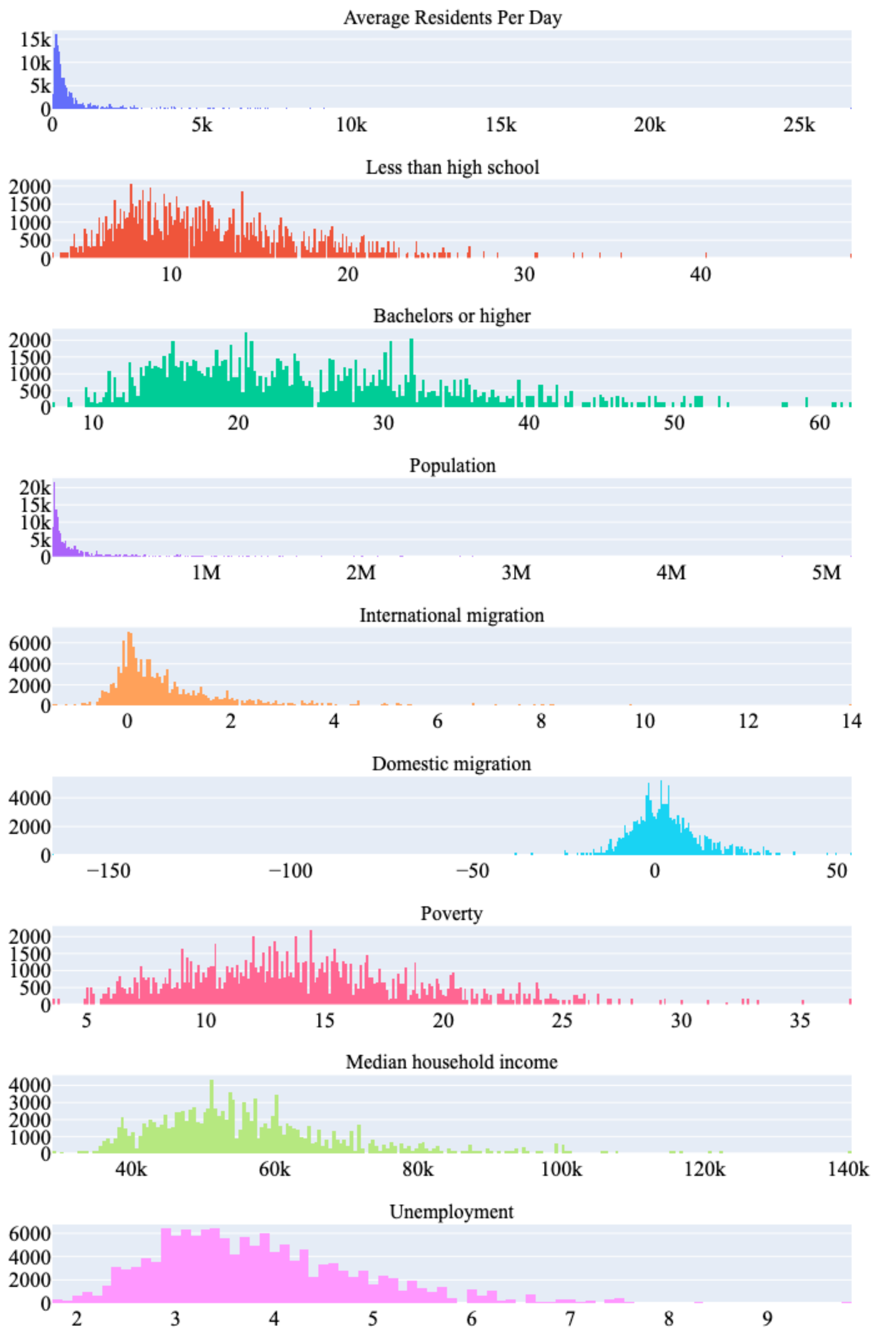


Figure 15: Histogram of control variables across sample counties

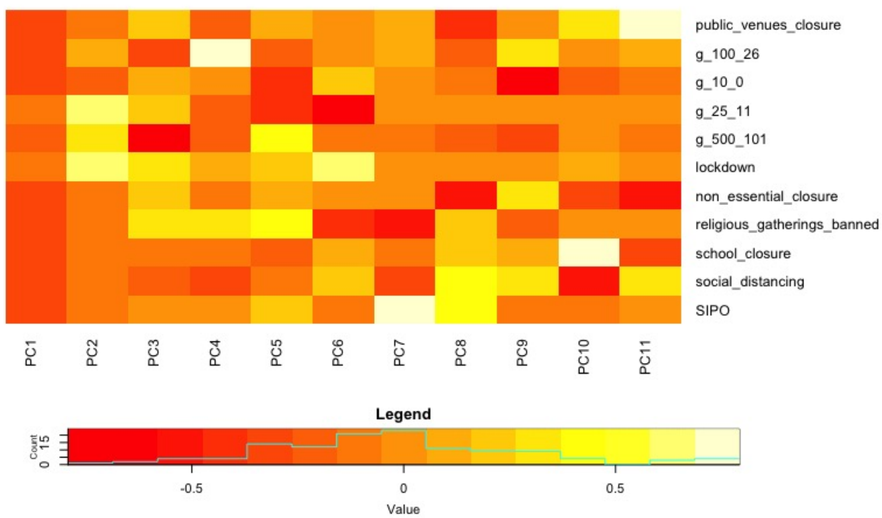


Figure 16: Loadings of various NPIs on the principal components