Prospective testing for the prevalence or transience of a shock effect over a set time horizon before it occurs

Abstract

We develop a hypothesis testing procedure to prospectively test whether an anticipated shock is likely to be transient or permanent over a set time horizon. We achieve this by borrowing knowledge from other time series that have undergone similar shocks for which post-shock outcomes are observed. These additional time series form a donor pool. For each time series in the donor pool we calculate a p-value corresponding to a hypothesis test on the relevance of the inclusion of shock-effect information in predicting the response over the time horizon. These p-values are then combined to form an aggregated p-value which guides one decision in determining whether the shock effect for the time series under study is expected to be prevalent or transient. This p-value can be computed before the shock-effect is observed in the time series under study provided one can form a suitable donor pool. Two real data examples for forecasting daily Conoco Phillips stock prices and monthly employment data as well as several simulations are provided for verification and illustration.

1 Introduction

In this article we provide forecasting methodology for assessing the transient or permanent behavior of an anticipated structural shock that has yet to have a measured impact on a time series under study. We focus on the setting in which a structural shock has occurred and one desires predictions to be made before post-shock responses are observed over a set time horizon H. Specific interest is in determining whether the shock is expected to be permanent or transient over H.

This is a general problem facing many real life applications. For example, one may acquire terrible or great news about a company and desire to determine whether that news is bound to impact the stock price of that company over a relevant time period. Companies may be interested in forecasting the demand of their products after they were involved in a brand crisis, but they only have recent sales data from pre-crises times. Policy makers, economists, and citizens alike may be interested in the expected future behavior of the unemployment rate in the upcoming quarter or year. Standard forecasting methods may not yield any guidance on the post-shock trajectories [Baumeister and Kilian, 2014b]. That being said, one may be able to construct credible forecasts in this setting [Lin and Eck, 2021]. The procedure of Lin and Eck [2021] works by supplementing the present forecast with past data borrowed from other time series which contain post-shock trajectories arising from materially similar structural shocks.

The method of Lin and Eck [2021] showed success in the nowcasting setting in which one is interested in predicting the value of the first post-shock response. As it currently stands this method is not appropriate for more nuanced forecasts involving h-ahead time points. In this article we develop a hypothesis testing procedure for determining whether or not time series which incorporate shock information exhibit superior predictive performance over forecasts which do not incorporate such information over a set time horizon. This effectively tests whether a shock is transient or permanent over a set time horizon.

The core idea of our methodology is to first form a donor pool of time series that have undergone similar shocks, then perform the multi-horizon forecast testing procedure of Quaedvlieg [2021] with respect to each of these donors to determine whether the shock is prevalent or transient, and then aggregate the decisions from the donors to make a final decision of transience or prevalence for the time series under

study for which no post-shock observations are observed. The multi-horizon forecast testing procedure of Quaedvlieg [2021] is a natural extension of Diebold and Mariano [1995]. The post-shock setting follows similar logic as that in Lin and Eck [2021]. Our procedure has conceptual similarities with conditional forecasting [Baumeister and Kilian, 2014b, Kilian and Lütkepohl, 2017], time series pooling using cross-sectional panel data [Ramaswamy et al., 1993, Pesaran et al., 1999, Hoogstrate et al., 2000, Baltagi, 2008, Koop and Korobilis, 2012, Liu et al., 2020], forecasting with judgement and models [Svensson, 2005, Monti, 2008], synthetic control methodology [Abadie et al., 2010], synthetic interventions [Agarwal et al., 2020b,a], and expectation shocks [Croushore and Evans, 2006, Baumeister and Kilian, 2014a, Clements et al., 2019]. The synthetic intervention approach in Agarwal et al. [2020a] is particularly relevant to our method. These authors construct a panel of countries undergoing a variety of interventions and use it to estimate what COVID case trajectories for a particular country would have looked like under different chosen courses of action. They then demonstrate that the method well replicates the observed trajectories of countries under the chosen intervention. This method works well because there are multiple countries undergoing similar interventions in response to a common shock which occurs at a similar past time for each country. In our setting, we do not specify that the time series in the donor pool be temporally similar to the time series under study, but we require that the shocks to those time series be similar to the shock of the time series under study. In Lin and Eck [2021] this similarity was stated in the form of a shock-effect distribution. In the present work, we parameterize similarity into post-shock dynamics for an autoregressive process.

As alluded to above, the success of our approach hinges on the construction of a suitable donor pool that possess similar post-shock dynamics to the time series under study. We develop a pragmatic parametric time series model which parameterizes such a structure. Without such a parameterization there is no avenue for connecting decisions made regarding the post-shock information in the donor pool with the yet-to-be observed post-shock information corresponding to the time series under study. The parameterization that we use states that the post-shock dynamics are determined in part by observed covariates collected right before the first post-shock response is observed. We demonstrate the utility of our prevalence testing approach on two real data examples forecasting daily Conoco Phillips stock prices and monthly employment data as well as several simulation examples. Our simulation examples show settings in which our method is expected to work well and demonstrate that the appropriateness of the method decreases as the parametric specifications break down. In the real data examples, we show that the decisions obtained from the donor pool match those that would be observed in the future.

2 Setting

We should make a picture and/or create an algorithm that communicates how our method works in a simple setting

We will suppose that a researcher has multivariate time series data $\mathbf{y}_{i,t}$, $t = 1, \ldots, T_i$ and $i = 1, \ldots, n+1$. We let $\mathbf{y}_{i,t} = (y_{i,t}, \mathbf{x}_{i,t})$ where $y_{i,t}$ is a scalar response and $\mathbf{x}_{i,t}$ is a vector of covariates that are revealed to the analyst prior to the observation of $y_{1,t}$. We will suppose that the analyst in interesting in forecasts over a time horizon H where $H > 1$ to define an interesting problem that differs from Lin and Eck [2021]. We will suppose that each time series $\mathbf{y}_{i,t}$ undergoes a shock at time $T_i^* \leq T_i + H$ for $i \geq 2$, and we will suppose that $T_1^* = T_1 + 1$. We will specify that $\mathbf{x}_{i,t=T_i^*}$ is observed before the shock takes effect on $y_{i,t=T_i^*}$ for all i. Under this setup there is no post-shock information for the series $y_{1,t}$.

In this article our goal will be to leverage the post-shock information in the donor pool of series $y_{i,t}$, $i = 2, \ldots, n+1$ to test whether the anticipated shock is expected to be transient or persistent over the time horizon H for the time series under study indexed by $i = 1$. The methodology developed within will focus on the comparison of two forecasts for $y_{i,t}^h$ at multiple horizons, $h = 1, \ldots, H$ where one forecast incorporates the shock effect and the other does not. We define these forecasts for each $i = 1, \ldots, n+1$,

$t = 1, \ldots, T_i$ and $h = 1, \ldots, H$ as

$$\hat{y}_{i,t}^{1,h} \text{ and } \hat{y}_{i,t}^{2,h},$$

where $y_{i,t}^{1,h}$ is the forecast for $y_{i,t}$ that accounts for the yet-to-be observed shock and is based on the information set $\mathcal{F}_{t-h}$, and $\hat{y}_{i,t}^{2,h}$ is defined similarly for the forecast that does not include any shock effect information.

Our methodology involves first comparing the performance of $\hat{y}_{i,t}^{1,h}$ and $\hat{y}_{i,t}^{2,h}$ for each time series in the donor pool ($i = 2, \ldots, n+1$). We then aggregate the decisions made in the donor pool by the similarity of the underlying time series to the time series under study. The first step in the process employs the recent multi-horizon forecast comparison methodology of Quaedvlieg [2021] for comparing the predictive abilitiy of forecasts jointly across all horizons of a forecast path, $h = 1, \ldots, H$. The second steps involves aggregation techniques motivated by the supposed data-generating process. These aggregation strategies are motivated by the post-shock aggregation techniques in Lin and Eck [2021]. We then report whether an anticipated shock is expected to be transient over the horizon H by using an aggregated decision from the donor pool.

(this needs work) In Section 3.1 we motivate time series models that we consider. In Section 3.2 we discuss the background in Quaedvlieg [2021] that is necessary for our post-shock testing procedure. In Section 3.3 we present aggregation strategies that we consider which inform our decision about the expected shock behavior. In Section 3.4 we provide a heuristic setting that demonstrates when our proposed procedure is expected to work well.

3 Methodological details

3.1 Model setup

We now describe the assumed autoregressive models with random effects for which post-shock aggregated estimators are provided. The model $\mathcal{M}$ is defined as

$$\mathcal{M}: \begin{aligned} y_{i,t} &= \eta_i + \left(\textstyle\sum_{j=1}^{q_1} \phi_{i,j} y_{i,t-j} + \sum_{j=0}^{q_2-1} \theta_{i,j+1}' \mathbf{x}_{i,t-j}\right)(1 - D_{i,t}) + f(\mathcal{F}_{i,t}, \alpha_i) D_{i,t} + \varepsilon_{i,t}, \\ f(\mathcal{F}_{i,t}, \alpha_i) &= \alpha_i + \textstyle\sum_{j=1}^{q_1} \tilde{\phi}_{i,j} y_{i,t-j} + \sum_{j=0}^{q_2-1} \tilde{\theta}_{i,j+1}' \mathbf{x}_{i,t-j}, \\ \alpha_i &= \mu_\alpha + \delta_i' \mathbf{x}_{i,T_i^*+1} + \varepsilon_{\alpha,i}, \end{aligned} \tag{1}$$

where $D_{i,t} = I(t \geq T_i^* + 1)$, $I(\cdot)$ is the indicator function, $\mathbf{x}_{i,t} \in \mathbb{R}^p$ are fixed with $p \geq 1$. Let $\phi_i = (\phi_{i,1}, \ldots, \phi_{i,q_1})'$, $\theta_i = (\theta_{i,1}, \ldots, \theta_{i,q_2})'$, $\tilde{\phi}_i = (\tilde{\phi}_{i,1}, \ldots, \tilde{\phi}_{i,q_1})'$, $\tilde{\theta}_i = (\tilde{\theta}_{i,1}, \ldots, \tilde{\theta}_{i,q_2})'$, $\delta_i = (\delta_{i,1}, \ldots, \delta_{i,p})'$, and suppose that the regression coefficients in (1) have the following hierarchical random effects structure:

$$\begin{aligned} \eta_i &\overset{iid}{\sim} \mathcal{F}_\eta \text{ with } \mathrm{E}_{\mathcal{F}_\eta}(\eta_i) = \mu_\eta, \mathrm{Var}_{\mathcal{F}_\eta}(\eta_i) = \sigma_\eta^2, \\ \phi_{i,j} &\overset{iid}{\sim} \mathcal{F}_{\phi_j} \text{ where } |\phi_{i,j}| < 1, \\ \theta_{i,j} &\overset{iid}{\sim} \mathcal{F}_{\theta_j} \text{ with } \mathrm{E}_{\mathcal{F}_{\theta_j}}(\theta_{i,j}) = \mu_{\theta_j}, \mathrm{Var}_{\mathcal{F}_{\theta_j}}(\theta_{i,j}) = \Sigma_{\theta_j}^2, \\ \delta_i &\overset{iid}{\sim} \mathcal{F}_\delta \text{ with } \mathrm{E}_{\mathcal{F}_\delta}(\delta_i) = \mu_\delta, \mathrm{Var}_{\mathcal{F}_\delta}(\delta_i) = \Sigma_\delta, \\ \varepsilon_{i,t} &\overset{iid}{\sim} \mathcal{F}_{\varepsilon_i} \text{ with } \mathrm{E}_{\mathcal{F}_{\varepsilon_i}}(\varepsilon_{i,t}) = 0, \mathrm{Var}_{\mathcal{F}_{\varepsilon_i}}(\varepsilon_{i,t}) = \sigma_i^2, \\ \varepsilon_{\alpha,i} &\overset{iid}{\sim} \mathcal{F}_{\varepsilon_\alpha} \text{ with } \mathrm{E}_{\mathcal{F}_{\varepsilon_\alpha}}(\varepsilon_{\alpha,i}) = 0, \mathrm{Var}_{\mathcal{F}_{\varepsilon_\alpha}}(\varepsilon_{\alpha,i}) = \sigma_\alpha^2, \\ \tilde{\phi}_{i,j} &\overset{ind}{\sim} \mathcal{F}_{\tilde{\phi}_j}(\mathbf{x}_{i,T_i^*+1}) \text{ where } |\tilde{\phi}_{i,j}| < 1, \\ \tilde{\theta}_{i,j} &\overset{ind}{\sim} \mathcal{F}_{\tilde{\theta}_j}(\mathbf{x}_{i,T_i^*+1}), \\ &\eta_i \perp\!\!\!\perp \phi_{i,j} \perp\!\!\!\perp \theta_{i,j} \perp\!\!\!\perp \tilde{\phi}_{i,j} \perp\!\!\!\perp \lambda_{i,j} \perp\!\!\!\perp \delta_i \perp\!\!\!\perp \varepsilon_{i,t} \perp\!\!\!\perp \tilde{\varepsilon}_i, \end{aligned} \tag{2}$$

where the distributions $\mathcal{F}_{\tilde{\phi}_j}(\mathbf{x}_{i,T_i^*+1})$ and $\mathcal{F}_{\tilde{\theta}_{i'}}(\mathbf{x}_{i,T_i^*+1})$ satisfy $\|\mathcal{F}_{\tilde{\phi}_j}(\mathbf{x}_{i,T_i^*+1}) - \mathcal{F}_{\tilde{\phi}_j}(\mathbf{x}_{i',T_i^*+1})\|_p \to 0$ and $\|\mathcal{F}_{\tilde{\theta}_j}(\mathbf{x}_{i,T_i^*+1}) - \mathcal{F}_{\tilde{\theta}_j}(\mathbf{x}_{i',T_i^*+1})\|_p \to 0$ as $\|\mathbf{x}_{i,T_i^*+1} - \mathbf{x}_{i',T_i^*+1}\| \to 0$ where $\|\cdot\|_p$ is a distance metric for distributions and $\|\cdot\|$ is a norm for elements in Euclidean space. Note that for model (1) to be of use for post-shock forecasting, the variation in $\mathcal{F}_{\tilde{\phi}}$ and $\mathcal{F}_{\tilde{\theta}}$ must be small relative to the signal strength which is captured in $\mathbf{x}_{i,T_i^*+1}$.

We see that model (1) with its accompanying random effects structure (2) is flexible enough to capture changing structural dynamics as well as a mean-shift, and is a generalization of the models mentioned in Lin and Eck [2021]. Note that the flexibility of allowing for changing structural dynamics is useful for fitting purposes, but it does not necessarily reflect the state of the world in the post-shock setting. It is of course possible that there is no shock or a simple mean-shift. In fact, the simple mean shift model can be fit and one can test whether the additional flexibility of model (1) is warranted for series in the donor pool. A general testing procedure is discussed in Section 3.3 (need to discuss that model selection is aggregated, the primary goal concerns the series under study).

The post-shock dynamic changes depend heavily on the value of $\mathbf{x}_{i,T_i^*+1}$, the covariates recorded right before the first post-shock response is observed. Under this setup any two series i, j with small $\|\mathbf{x}_{i,T_i^*+1} - \mathbf{x}_{j,T_i^*+1}\|$ are expected to experience similar structural changes. This makes distance-based weighting an attractive avenue.

Next, we define the h-step forecast for $\mathcal{M}$ which plays an essential role in the framework of Quaedvlieg [2021]. Denote h-step forecast for $y_{i,t}$ by $\hat{y}_{i,t}^{1,h}$. Let the ordinary least squares (OLS) estimate for

$$\{(\eta_i, \alpha_i, \phi_{i,j_1}, \boldsymbol{\theta}'_{i,j_2}, \tilde{\phi}_{i,j_1}, \tilde{\boldsymbol{\theta}}'_{i,j_2})' \colon i = 2, \ldots, n+1, j_1 = 1, \ldots, q_1, j_2 = 1, \ldots, q_2\}$$

be

$$\{(\hat{\eta}_i, \hat{\alpha}_i, \hat{\phi}_{i,j_1}, \hat{\boldsymbol{\theta}}'_{i,j_2}, \hat{\tilde{\phi}}_{i,j_1}, \hat{\tilde{\boldsymbol{\theta}}}'_{i,j_2})' \colon i = 2, \ldots, n+1, j_1 = 1, \ldots, q_1, j_2 = 1, \ldots, q_2\}.$$

Based on the training data $\{(\mathbf{x}_{i,t}, y_{i,t}) \colon i = 2, \ldots, n+1, t = 1, \ldots, T_i\}$, for $h = 1, \ldots, H$, the h-step forecast $\hat{y}_{i,t}^{1,h}$ is defined as

$$\hat{y}_{i,t}^{1,h} = \hat{\eta}_i + \left(\sum_{j=1}^{q_1} \hat{\phi}_{i,j} \hat{y}_{i,t-j}^{1,h-j} + \sum_{j=0}^{q_2-1} \mathbf{x}'_{i,t-j} \hat{\boldsymbol{\theta}}_{i,j+1} \right) (1 - D_{i,t}) + f(\hat{\mathcal{F}}_{i,t}, \hat{\alpha}_i) D_{i,t},$$

$$f(\hat{\mathcal{F}}_{i,t}, \hat{\alpha}_i) = \hat{\alpha}_i + \sum_{j=1}^{q_1} \hat{\tilde{\phi}}_{i,j} \hat{y}_{i,t-j}^{1,h-j} + \sum_{j=0}^{q_2-1} \mathbf{x}'_{i,t-j} \hat{\tilde{\boldsymbol{\theta}}}_{i,j+1},$$

where it is allowed that $t > T_i$, and we set $\hat{y}_{i,t}^{s} = y_{i,t}$ for $t \leq T_i$ and $s \leq 0$. The h-step forecast corresponding to the model that does include any post-shock changes $\hat{y}_{i,t}^{2,h}$ is defined similarly,

$$\hat{y}_{i,t}^{2,h} = \hat{\eta}_i + \left(\sum_{j=1}^{q_1} \hat{\phi}_{i,j} \hat{y}_{i,t-j}^{2,h-j} + \sum_{j=0}^{q_2-1} \mathbf{x}'_{i,t-j} \hat{\boldsymbol{\theta}}_{i,j+1} \right).$$

3.2 Multi-horizon forecasting procedure

In this section we discuss the background of the multi-horizon forecasting procedure [Quaedvlieg, 2021] that is necessary for the investigation of shock behavior in our anticipated post-shock setting. The initial building blocks in the assessment of predictive ability of forecasts $\hat{y}_{i,t}^{1,h}$ and $\hat{y}_{i,t}^{2,h}$ will make use of a loss differential

$$\mathbf{d}_{i,t} = \mathbf{L}_{i,t,2} - \mathbf{L}_{i,t,1},$$

where $\mathbf{L}_{i,t,j} \in \mathbb{R}^H$ has elements $L^h(y_{i,t}, \hat{y}^{j,h}_{i,t})$, $j = 1, 2$, and L is a loss function. Define the expected loss differential as $E(\mathbf{d}_{i,t}) = \mu_{i,t}$ and let the average expected loss differential be defined as $\mu_i = \lim_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} \mu_{i,t}$. We will then judge forecasts $\hat{y}^{1,h}_{i,t}$ and $\hat{y}^{2,h}_{i,t}$ based on the average superior predictive ability (aSPA) [Quaedvlieg, 2021]. The aSPA investigates forecast comparisons based on their weighted average loss difference

$$\mu_i^{(\text{Avg})} = \mathbf{w}_i'\mu_i = \sum_{h=1}^{H} w_{i,h}\mu_i^h,$$

with weights $\mathbf{w}_i$ that sum to one. Note that aSPA requires the user to take a stand on the relative importance of under-performance at one horizon against out-performance at another, and note that it is likely that $\mu_i^h > 0$ for h closer to 1 since the user expects that a structural shock will occur and the structural shock is taken into account by forecast 1. With this caveat in mind we will use $w_{i,h} = 1/H$, and we will consider different time horizons H in some of our examples.

Define $\mathbf{D}_i = (\mathbf{d}'_{i,T_i^*+H+1}, \ldots, \mathbf{d}'_{i,T_i})$, which is the loss differential matrix of ith time series. Note that $\mathbf{D}_i$ is of dimension $T_i - T_i^* - H$. $\mathbf{D}_i$ compares h-step forecasts of $y_{i,t}$ for $t = T^* + H + 1$ to $t = T$ and $h = 1, \ldots, H$. In this design, adjusted and unadjusted forecasts are different with probability one such that elements of $\mathbf{D}_i$ are different from zero with probability one. This prevents excessive zeroes to affect our inference results if $t = H+K+1, \ldots, T$, where K is the training sample size for parameter estimation of h-step forecasts.

We consider a simple test for average SPA, based on the weighted-average loss differential. The associated null is

$$H^0_{i,\text{aSPA}} : \mu_i^{(\text{Avg})} \leq 0. \tag{3}$$

A studentized test statistic corresponding to the null hypothesis (3) is of the form

$$t_{i,\text{aSPA}} = \frac{\sqrt{T}_i \bar{d}_i}{\hat{\zeta}_i}, \tag{4}$$

where $\bar{d}_i = \mathbf{w}_i'\mathbf{d}_i$ and we choose to estimate $\zeta_i = \sqrt{\mathbf{w}_i'\Omega_i\mathbf{w}_i}$ directly based on $\mathbf{w}_i'\mathbf{d}_{i,t}$ using the HAC estimator [Giacomini and White, 2006] where $\Omega_i = \text{avar}\left(\sqrt{T}_i(\bar{d}_i - \mu_i)\right)$.

A p-value corresponding to the hypothesis test (3) is computed using a moving block bootstrap (MBB) of Künsch [1989] and Liu and Singh [1992]. In the moving block bootstrap (MBB), a pseudo time-series is constructed by resampling blocks of length ℓ from the original data.

Let $l = o(T^{1/2})$ be the block length of the MBB where we assume that $T_i = lU_i$. Let $I_1, \ldots, I_{K_i}$ be iid random variables uniformly distributed on $\{1, \ldots, T_i - l + 1\}$, and define the array

$$\tau_{T_i} = \{I_1 + 1, \ldots, I_1 + l, \ldots, I_{K_i} + 1, \ldots, I_{K_i} + l\}.$$

The pseudo time-series is therefore $\mathbf{d}^b_{i,t} = \mathbf{d}^b_{i,\tau_{T_i}}$, with elements $d^{hb}_{i,t}$.

$$\left(\hat{\omega}_i^{hb}\right)^2 = \frac{1}{U_i}\sum_{k=1}^{U_i}\left[\frac{1}{l}\left(\sum_{t=1}^{l} d^{hb}_{i,(k-1)l+t} - \bar{d}_i^{hb}\right)^2\right], \tag{5}$$

where $\bar{d}_i^{hb} = \frac{1}{T_i}\sum_{t=1}^{T_i} d^{hb}_{i,t}$. From the conditions of Theorem 1 and Corollary 1 in Quaedvlieg [2021] we have

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}^b\left[\sqrt{T_i}\frac{\mathbf{w}_i'\bar{\mathbf{d}}_i^b - \mathbf{w}_i'\bar{\mathbf{d}}_i}{\hat{\zeta}_i^b}\right] - \mathbb{P}\left[\sqrt{T_i}\frac{\mathbf{w}_i'\bar{\mathbf{d}}_i - \mathbf{w}_i'\mu_i}{\hat{\zeta}_i^b}\right]\right| \xrightarrow{P} 0, \tag{6}$$

as $T_i \to \infty$ where $l = l_{T_i} = o(\sqrt{T_i})$. We can compute a bootstrap p-value as

$$\hat{p}_i^B = \frac{1}{B}\sum_{b=1}^{B} 1\{t_{i,\text{aSPA}} < t^b_{i,\text{aSPA}}\}. \tag{7}$$

The results (6) implies that

$$|\hat{p}_i^B - \hat{p}_i| \xrightarrow{P} 0, \quad \text{as} \quad T_i, B \to \infty,$$

where $\hat{p}_i = \mathbb{P}(t_{i,\alpha} < t_{i,\text{aSPA}})$ and $t_{i,\alpha}$ is a critical value corresponding to the distribution of $t_{i,\text{aSPA}}$. Therefore bootstrap inference in the donor pool is a good approximation.

3.3 Aggregation methods

In this section we discuss aggregation techniques for combining the multi-horizon forecast p-values in Section 3.2 computed from forecasts from models in Section 3.1 to illicit a decision about the transient or persistence of a structural shock.

The post-shock dynamics in model (1) with random effect structure (2) are explicitly constructed so that series with similar covariates at time $T_i^* + 1$ have similar post-shock dynamics. This facilitates the use of simple p-value combination methods provided that there exists donors i in the donor pool which have $\mathbf{x}_{i,T_i^*+1}$ close to $\mathbf{x}_{1,T_1^*+1}$. We will motivate two such p-value combination approaches.

The first approach is a weighted average of the p-values computed in the donor pool which are calculated as in Section 3.2. The weights reflect the similarity between $\mathbf{x}_{i,T_i^*+1}$ and $\mathbf{x}_{1,T_1^*+1}$, and are similar to the weights computed in Lin and Eck [2021]. We now discuss the details. Let

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}'_{2,T_2^*+1} \\ \vdots \\ \boldsymbol{x}'_{n+1,T_{n+1}^*+1} \end{pmatrix}.$$

In this paper, we consider this simple form of $\boldsymbol{X}$ for illustrative purpose. However, more general $\boldsymbol{X}$'s are allowable. For example, $\boldsymbol{X}$ can also include $\boldsymbol{x}_{i,T_i^*-j+1}$ for $j = 2, \ldots, q_2$, which considers the lagged covariates in the dynamic model (1). Let $\boldsymbol{X}^*$ be the same as $\boldsymbol{X}$ with each column scaled to have mean 0 and standard deviation 1. We compute the weights $\mathbf{W}^*$ based on $\boldsymbol{X}^*$ according to procedures detailed in Section 2.2 of Lin and Eck [2021]. Note that the modeling part for the response $y_{i,t}$ does not necessarily involve scaled covariates. From this construction we obtain the weighted average p-value

$$\hat{p}_1 = \sum_{i=2}^{n+1} w_i^* \hat{p}_i^B, \tag{8}$$

where $\hat{p}_i^B$ is given as (7) and w_i^* is the weight corresponding to the similarity between the first and $(i+1)$th row of $\mathbf{X}^*$.

Need to make a decision about voting and weighted voting; not sure if it is more robust or efficient than the weighted p-value approach. If it is not superior than the weighted approach in any way then we should remove it. We now motivate a voting method and a weighted voting method based on the p-values computed in the donor pool. Given a significance level α and estimates $\{\hat{p}_i^B : i = 2, \ldots, n+1\}$, we compute the vote

$$V = I\Big\{\frac{1}{n}\sum_{i=2}^{n+1} I(\hat{p}_i^B \le \alpha) \ge 0.5\Big\},$$

where $I(\cdot)$ is an indicator function. The idea of voting is very intuitive in the sense that donors vote for the shock persistence or transience using a hypothesis testing approach and the one receiving more votes is returned. Second, the weighted voting generalizes this idea by making using of the similarity weights $\mathbf{W}^*$. It is defined as

$$V^* = I\Big\{\sum_{i=2}^{n+1} w_i^* I(\hat{p}_i^B \le \alpha) \ge 0.5\Big\}.$$

Algorithm 1 describes the mechanics of computing $\mathbf{W}^*, \hat{p}_1, V, V^*$ in more detail. There are a couple of details that are worth noting. First, for $t = T_i + 1$ and $i \geq 2$, we used to the methodology of Lin and Eck [2021] to compute point forecast for $\hat{y}_{i,t}^{2,h}$. Note that for $t = T_i + 1$, using ordinary least square conditional forecast yields $\hat{y}_{i,t}^{2,h} = \hat{y}_{i,t}^{1,h}$ as those two forecast just differ by shock information, which is not available at this time point. We make such adjustment for the reason that the method of Lin and Eck [2021] is generally more likely to perform better than the ordinary forecast under certain conditions. Second, notice that we evaluate forecast comparison for $t = T_i^* + \lceil 1.5 \cdot \sqrt{T_i - T_i^*} \rceil + 1, \ldots, T_i$ with time length smaller than T_i. It is because the framework of Quaedvlieg [2021] needs training sample, which implicitly requires users to feed the algorithm with moderately large dataset for a more credible analysis.

Algorithm 1: Algorithm for computing $\hat{p}_1$, weighted voting, and $\mathbf{W}^*$

Input: B, ℓ – bootstrap sample size, the block size
γ – bandwidth parameter in heteroskedasticity and autocorrelation consistent covariance matrix estimation.
α – significance level
$\{(y_{i,t}, \mathbf{x}_{i,t}) \colon i = 2, \ldots, n+1, t = 0, \ldots, T_i\}$ – the data
$\{T_i^* \colon i = 1, \ldots, n+1\}$ – the time point just before the shock
$\{K_i \colon i = 2, \ldots, n+1\}$ – training sample size for h-step forecast for different time series
q_1, q_2 – user-specified parameters in $\mathcal{M}$
H – maximum horizon

Result: (1) synthetic weights $\mathbf{W}^*$, (2) $\hat{p}_1$, and (3) decision based on voting and weighted voting

for $i = 2 : (n+1)$ **do**
  if $i > 2$ **then**
    Compute $\hat{\alpha}_{\text{wadj}}$ based on Lin and Eck [2021] with $\{T_j^* \colon j = 2, \ldots, i-1\}$ and $\{(y_{j,t}, \mathbf{x}_{j,t}) \colon j = 2, \ldots, i-1, t = 0, \ldots, T_j\}$
  end
  for $h = 1, \ldots, H$ **do**
    for $t = T_i^* + \lceil 1.5 \cdot \sqrt{T_i - T_i^*} \rceil + 1, \ldots, T_i$ **do**
      Compute $\hat{y}_{i,t}^{1,h}$ and $\hat{y}_{i,t}^{2,h}$ based on OLS estimation of $\mathcal{M}$ with $\{T_i^* \colon i = 1, \ldots, n+1\}$ and training data $\{(y_{i,t}, \mathbf{x}_{i,t}) \colon t = t - h - K_i + 1, \ldots, t - h\}$, where we further need the set of lagged response $\{y_{i,t} \colon t = t - h - K_i - \max\{q_1, q_2 - 1\} + 1, t - h - K_i\}$.
      Compute the squared error loss $l_{i,t,h,1}$ based on $y_{i,t}$ and $\hat{y}_{i,t}^{1,h}$
      Compute the squared error loss $l_{i,t,h,2}$ based on $y_{i,t}$ and $\hat{y}_{i,t}^{2,h}$
    end
  end
  Construct the loss differential matrix with element $l_{i,t,h,2} - l_{i,t,h,1}$ at tth row and hth column
  Conduct multiple horizon forecast comparison test of Quaedvlieg [2021] on the loss differential matrix $\{l_{i,t,h,2} - l_{i,t,h,1}\}_{t,h}$ with parameters ℓ, B, γ to obtain forecast comparison p-value $\hat{p}_i^B$, where the test uses moving block bootstrap with B and γ.
end
Compute synthetic weights $\mathbf{W}^* = (w_2^*, \ldots, w_{n+1}^*)$ based on $\{\mathbf{x}_{i,T_i^*+1} \colon i = 1, \ldots, n+1\}$
Compute the vote $V = I\Big\{n^{-1} \sum_{i=2}^{n+1} I(\hat{p}_i \leq \alpha) \geq 0.5\Big\}$
Compute the weighted vote $V^* = I\Big\{\sum_{i=2}^{n+1} w_i^* I(\hat{p}_i \leq \alpha) \geq 0.5\Big\}$

Proposition 1. *Suppose $p_2, \ldots, p_{n+1}$ is a sequence of pairwise independent p-values with $\mathbb{P}(p_i \leq \alpha) = \kappa$ for $i = 1, \ldots, n+1$, where α is the significance level, κ is a real-valued constant in $[0, 1]$, and p_1 is the*

p-value of the time series of interest. If $\mathbb{P}(p_1 \leq \alpha) \neq 0.5$, *as* $n \to \infty$,

$$\mathbb{E}\left\{ \left| I\left\{ \frac{1}{n}\sum_{i=2}^{n+1} I(p_i \leq \alpha) \geq 0.5 \right\} - I(p_1 \leq \alpha) \right| \right\} \to \begin{cases} 1 - \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) > 0.5 \\ \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) < 0.5 \end{cases}$$

Proposition 2. *Let* $(w_2, \ldots, w_{n+1})$ *be weights such that* $w_i \in [0,1]$ *and* $\sum_{i=2}^{n+1} w_i = 1$. *Define*

$$\mathcal{I}_n = \{i = 2, \ldots, n+1 \colon 0 < w_i < 1\}.$$

Suppose that $\mathcal{I}_n$ *is non-empty,* $|\mathcal{I}_n| \to \infty$ *as* $n \to \infty$, *and* $w_i b_n \leq K$ *for* $i \in \mathcal{I}_n$ *and some* $K > 0$, *where* $b_n \geq 1$ *and* $b_n \to \infty$ *as* $n \to \infty$. *Assume for* $i \in \mathcal{I}_n$, p_i *are pairwise independent p-values with*

$$\sum_{i \in \mathcal{I}_n} \mathbb{P}(p_i \leq \alpha) = \sum_{i \in \mathcal{I}_n} w_i \kappa_i \to \kappa_1,$$

where $\kappa_i = \mathbb{P}(p_i \leq \alpha) \in [0,1]$ *for* $i \in \mathcal{I}_n$, α *is the significance level,* $\kappa_1 = \mathbb{P}(p_1 \leq \alpha)$, *and* p_1 *is the p-value of the time series of interest. If* $\mathbb{P}(p_1 \leq \alpha) \neq 0.5$, *as* $n \to \infty$,

$$\mathbb{E}\left\{ \left| I\left\{ \sum_{i \in \mathcal{I}_n} w_i I(p_i \leq \alpha) \geq 0.5 \right\} - I(p_1 \leq \alpha) \right| \right\} \to \begin{cases} 1 - \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) > 0.5, \\ \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) < 0.5. \end{cases}$$

Note that asymptotic approximation for the misclassification error does not work when the weight is concentrated, i.e., $\mathcal{I}_n$ is finite even if $n \to \infty$, where $\mathcal{I}_n = \{i = 2, \ldots, n+1 \colon 0 < w_i < 1\}$. The intuition is that it is impossible to find a weight to constrain the random behavior of the Bernoulli random variables $I(p_i \leq \alpha)$ for $i \in \mathcal{I}_n$ such that the weak law of large number can apply. Nevertheless, in this case, if $\kappa_i = \mathbb{P}(p_i \leq \alpha)$ for $i \in \mathcal{I}_n$ is small, the approximation for the misclassification in Proposition 2 is favorable. This situation occurs very frequently in practice. It can be shown as below.

Note that the mean squared error of estimating $\mathbb{P}(p_1 \leq \alpha) = \kappa_1$ is

$$\begin{aligned} \mathbb{E}\left\{ \left\{ \sum_{i \in \mathcal{I}_n} w_i I(p_i \leq \alpha) - \kappa_1 \right\}^2 \right\} &= \left\{ \sum_{i \in \mathcal{I}_n} w_i \kappa_i - \kappa_1 \right\}^2 + \mathbf{Var}\left\{ \sum_{i \in \mathcal{I}_n} w_i I(p_i \leq \alpha) \right\} \\ &= \left\{ \sum_{i \in \mathcal{I}_n} w_i \kappa_i - \kappa_1 \right\}^2 + \sum_{i \in \mathcal{I}_n} w_i^2 \kappa_i (1 - \kappa_i) \\ &\overset{n \to \infty}{\to} \sum_{i \in \mathcal{I}_n} w_i^2 \kappa_i (1 - \kappa_i), \end{aligned}$$

where the last second step assumes that $\{I(p_i \leq \alpha) \colon i \in \mathcal{I}_n\}$ are independent and the last step uses the condition $\sum_{i \in \mathcal{I}_n} w_i \kappa_i \to \kappa_1$ as $n \to \infty$, which is used in Proposition 2. Notice that if κ_i is very small, the mean squared error of $\sum_{i \in \mathcal{I}_n} w_i I(p_i \leq \alpha)$ turns out to be small such that by the rationale of the proof of Proposition 1, the asymptotic approximation for the misclassification error in Proposition 2 should be acceptable.

We now motivate a model selection framework for comparing different models in the donor pool with a goal of inferring which model is most appropriate for the time series under study. Suppose we are interested in comparing two models $\mathcal{M}_0$ and $\mathcal{M}$ using AIC for forecasting the response in the time series of interest. However, direct comparison is not feasible as the data are not observed. Instead, we can conduct such inference based on the donor pool. Let the AIC of $\mathcal{M}_j$ for ith time series be a_i^j for $i = 1, \ldots, n+1$ and $j = 1, 2$. Let $\xi_i = \mathbb{P}(a_i^1 < a_i^2)$. Note that if the data generating process is based on $\mathcal{M}_0$ and is true for all the donors and time series of interest, ξ_i is expected to be high. Additionally, it is reasonable to assume $\xi_i = \xi$ for $\xi \in [0,1]$ and $i = 1, \ldots, n+1$. Notice that this setup is similar to the setup of weighted voting or voting we discussed before. So, under similar arguments, we propose the following corollary for synthetic pairwise model selection.

Corollary 1 (Synthetic pairwise model selection)**.** *Let* $(w_2, \ldots, w_{n+1})$ *be weights such that* $w_i \in [0,1]$ *and* $\sum_{i=2}^{n+1} w_i = 1$*. Define*

$$\mathcal{I}_n = \{i = 2, \ldots, n+1 \colon 0 < w_i < 1\}.$$

Suppose that $\mathcal{I}_n$ *is non-empty,* $|\mathcal{I}_n| \to \infty$ *as* $n \to \infty$*, and* $w_i b_n \leq K$ *for* $i \in \mathcal{I}_n$ *and some* $K > 0$*, where* $b_n \geq 1$ *and* $b_n \to \infty$ *as* $n \to \infty$*. Assume for* $i \in \mathcal{I}_n$*,* $\{a_i^1, a_i^2\}$ *are AICs of* $\mathcal{M}_0$ *and* $\mathcal{M}$ *for* i*th time series, which are pairwise independent of* $\{a_j^1, a_j^2\}$ *for* $j \neq i$ *and that*

$$\sum_{i \in \mathcal{I}_n} \mathbb{P}(a_i^1 < a_i^2) = \sum_{i \in \mathcal{I}_n} w_i \xi_i \to \xi_1,$$

where $\xi_i = \mathbb{P}(a_i^1 < a_i^2) \in [0,1]$ *for* $i \in \mathcal{I}_n$*. If* $\xi \neq 0.5$*, as* $n \to \infty$*,*

$$\mathbb{E}\left\{ \left| I\left\{ \sum_{i \in \mathcal{I}_n} w_i I(a_i^1 < a_i^2) \geq 0.5 \right\} - I(a_1^1 < a_1^2) \right| \right\} \to \begin{cases} 1 - \xi_1 & \text{if } \xi_1 > 0.5 \\ \xi_1 & \text{if } \xi_1 < 0.5 \end{cases}$$

Note that Corollary 1 gives an asymptotic upper bound for the misclassification rate of synthetic pairwise model selection, which is small if ξ_1 is large.

3.4 Heuristic setting demonstrating when prevalence testing works well

We now demonstrate an idealized theoretical setting for which our post-shock testing methodology involving a distance-based weighted average of p-values works well. The idea is that inferences based on distance-based weighted nearly recovers the time series under study when there exists a time series in the donor pool, indexed by i', such that $\mathbf{x}_{1,t} \approx \mathbf{x}_{i',t}$ and the signal strength in our modeling setup (1) and (2) is large relative to variation.

We first present some intermediary technical results and definitions. Let $\lambda_{\max,\theta_k}$, $\lambda_{\max,\theta_\delta}$, $\lambda_{\max,\gamma}$ respectively be the largest eigenvalues of Σ_{θ_k}, Σ_δ, and Σ_γ for $k = 1, \ldots, q_2$. Define

$$\begin{aligned}
\mu_{\min} &= \min\left(\mu_\eta, \mu_{\phi_j}, \mu_{\tilde{\phi}_j}, \min(\mu_\delta), \min(\mu_{\theta_k}), \min(\mu_{\tilde{\theta}_k}); j = 1, \ldots, q_1, k = 1, \ldots, q_2\right), \\
\mu_{\max} &= \max\left(\mu_\eta, \mu_{\phi_j}, \mu_{\tilde{\phi}_j}, \max(\mu_\delta), \max(\mu_{\theta_k}), \max(\mu_{\tilde{\theta}_k}); j = 1, \ldots, q_1, k = 1, \ldots, q_2\right), \\
\lambda_{\max} &= \max\left(\lambda_{\max,\theta_k}, \lambda_{\max,\theta_\delta}, \lambda_{\max,\gamma}; k = 1, \ldots, q_2\right), \\
\sigma^2_{\max} &= \max\left(\sigma^2_\eta, \sigma^2_{\phi_j}, \sigma^2_i, \sigma^2_\alpha, \sigma^2_{\tilde{\phi}_j}; i = 1, \ldots, n+1, j = 1, \ldots, q_1\right),
\end{aligned}$$

and let $\sigma^2 = \max(\lambda_{\max}, \sigma^2_{\max})$. Now fix i' where we suppose that $\|\mathbf{x}_{1,t} - \mathbf{x}_{i',t}\| < \epsilon_1$ for all t and some choice of ϵ_1. Pick some large $M > 0$. Fix $T = T_1 = T_{i'}$ and $T^* = T_1^* = T_{i'}^*$. Define the events

$$\begin{aligned}
E_1 :=& \Big\{ \max\left(|\|\delta_1 - \delta_{i'}\|, |\tilde{\phi}_{1,j} - \tilde{\phi}_{i',j}|, \|\tilde{\theta}_{1,k} - \tilde{\theta}_{i',k}\|, |\tilde{\varepsilon}_t| \right) < \epsilon_1 : \\
& \qquad j = 1, \ldots, q_1, k = 1, \ldots, q_2, t = 1, \ldots, T \Big\}, \\
E_2 :=& \{\|(\boldsymbol{\eta}', \boldsymbol{\phi}', \boldsymbol{\theta}', \boldsymbol{\delta}', \tilde{\boldsymbol{\phi}}', \tilde{\boldsymbol{\theta}}')'\|_\infty < M_1\},
\end{aligned}$$

where $\tilde{\varepsilon}_t = \varepsilon_{1,t} - \varepsilon_{i',t} + D_{i,t}(\varepsilon_{\alpha,1} - \varepsilon_{\alpha,i'})$ and $\boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\delta}, \tilde{\boldsymbol{\phi}}$, and $\tilde{\boldsymbol{\theta}}$ are vectors containing the random effects in (2) for indices 1 and i', and we take $\|\boldsymbol{a}\|_\infty = \max_j(|a_j|)$. We will also suppose that $\|\mathbf{x}_{j,t}\|_\infty < M_1$ for $j \in \{1, i'\}$ and all $t = 1, \ldots, T$. Note that both $\mathbb{P}(E_1^C) \to 0$ and $\mathbb{P}(E_2^C) \to 0$ as $\sigma^2 \to 0$ with T and n fixed. For simplicity we will set $q_1 = q_2 = 1$.

For all $t > T^*$ let $\tilde{y}_{i',t}$ possess the same dynamics as $y_{i',t}$ except that this process is shifted in a manner to reflect how $y_{i,t}$ changes for all $t > T^*$. More formally, we will let

$$\tilde{y}_{i',t} = \begin{cases} y_{i,t} & \text{when } t < T^* + 1; \\ \eta_1 + \mu_\alpha + \delta_{i'}' \mathbf{x}_{i',T^*+1} + \tilde{\phi}_{i',t} y_{1,t} + \tilde{\theta}_{i'}' \mathbf{x}_{i',t} + \varepsilon_{i',t} + \varepsilon_{\alpha,i'} & \text{when } t = T^* + 1; \\ \eta_1 + \mu_\alpha + \delta_{i'}' \mathbf{x}_{i',T^*+1} + \tilde{\phi}_{i',t} \tilde{y}_{i',t} + \tilde{\theta}_{i'}' \mathbf{x}_{i',t} + \varepsilon_{i',t} + \varepsilon_{\alpha,i'} & \text{when } t > T^* + 1. \end{cases}$$

The purpose of $\tilde{y}_{i',t}$ is to place the post-shock responses at the same starting point which allows us to avoid strong parametric specifications on the pre-shock dynamics. This series also contains terms which are estimable from the data. Define the event

$$E_3 := \{\max\left(y_{1,t}, \tilde{y}_{i',t}\right) < M_2 : t = 1, \ldots, T\},$$

where M_2 is chosen large enough so that $\mathbb{P}(E_3^c) \to 0$ as $\sigma^2 \to 0$. Let $M = \max(M_1, M_2)$. We now consider the recursion of $|y_{1,t} - \tilde{y}_{i',t}|$ for $t > T^*$ conditional on $E = E_1 \cup E_2 \cup E_3$,

$$\begin{aligned}
|y_{1,t} - \tilde{y}_{i',t}| &= |(\delta_1' \mathbf{x}_{1,T^*+1} - \delta_{i'}' \mathbf{x}_{i',T^*+1}) + (\tilde{\phi}_1 y_{1,t-1} - \tilde{\phi}_{i'} \tilde{y}_{i',t-1}) + (\tilde{\theta}_1' \mathbf{x}_{1,t} - \tilde{\theta}_{i'}' \mathbf{x}_{i',t}) + \tilde{\varepsilon}_t| \\
&\leq |\delta_1' \mathbf{x}_{1,T^*+1} - \delta_{i'}' \mathbf{x}_{i',T^*+1}| + |\tilde{\phi}_1 y_{1,t-1} - \tilde{\phi}_{i'} \tilde{y}_{i',t-1}| + |\tilde{\theta}_1' \mathbf{x}_{1,t} - \tilde{\theta}_{i'}' \mathbf{x}_{i',t})| + \epsilon_1 \\
&\leq (8M+1)\epsilon_1 + |\tilde{\phi}_1 - \tilde{\phi}_{i'}||y_{i',t-1}| + |\tilde{\phi}_1||y_{1,t-1} - \tilde{y}_{i',t-1}| \\
&\leq (8M+1)\epsilon_1 + |\tilde{\phi}_1 - \tilde{\phi}_{i'}||y_{i',t-1}| \\
&\quad + |\tilde{\phi}_1| \left\{(8M+1)\epsilon_1 + |\tilde{\phi}_1 - \tilde{\phi}_{i'}||\tilde{y}_{i',t-2}| + |\tilde{\phi}_1||y_{1,t-2} - \tilde{y}_{i',t-2}|\right\} \\
&= \epsilon_1(8M+1)(1 + |\tilde{\phi}_1|) + |\tilde{\phi}_1||\tilde{\phi}_1 - \tilde{\phi}_{i'}|(|\tilde{y}_{i',t-1}| + |y_{i',t-2}|) + |\tilde{\phi}_1|^2 |y_{1,t-2} - \tilde{y}_{i',t-2}| \\
&\leq \epsilon_1(8M+1) \sum_{j=0}^{t-T^*-1} |\tilde{\phi}_1|^j + |\tilde{\phi}_1 - \tilde{\phi}_{i'}| \sum_{j=1}^{t-T^*} |\tilde{\phi}_1|^{j-1} |\tilde{y}_{i',t-j}| + |\tilde{\phi}_1|^{t-T^*} |y_{1,T^*} - \tilde{y}_{i',T^*}| \\
&= \epsilon_1(8M+1) \sum_{j=0}^{t-T^*-1} |\tilde{\phi}_1|^j + |\tilde{\phi}_1 - \tilde{\phi}_{i'}| \sum_{j=1}^{t-T^*} |\tilde{\phi}_1|^{j-1} |\tilde{y}_{i',t-j}|.
\end{aligned}$$

With T set, we can choose ϵ_1 small enough so that $|y_{1,t} - \tilde{y}_{i',t}| < \epsilon_2$ for all $T^* < t \leq T$.

Now let $\epsilon_3 = \max(\epsilon_1, \epsilon_2)$ so that

$$|y_{1,t} - \tilde{y}_{i',t}| < \epsilon_3, \text{ and } \|\mathbf{x}_{1,t} - \mathbf{x}_{i',t}\| < \epsilon_3. \tag{9}$$

Let $\tilde{\beta}$ be the OLS solution of model parameters with respect to the $\tilde{y}_{i',t}$ series, and let $\hat{\beta}$ be the OLS solution of model parameters with respect to the $y_{1,t}$ series. On E, these estimates will be very close to each other. We use Theorem 2.7.2 in Golub and Van Loan [1996] to construct an explicit bound on $\|\tilde{\beta} - \hat{\beta}\|$. Following the notation of Section 2.7 in Golub and Van Loan [1996] we will let A be the design matrix with respect to the $y_{1,t}$ series so that $A\hat{\beta} = \mathbf{y}_1$ where $\mathbf{y}_1 = (y_{1,1}, \ldots, y_{1,T})'$, $\kappa(A) = \|A\|\|A^{-1}\|$ be the condition number of A where $\|\cdot\|$ is a matrix norm, and we will specify that $r = \epsilon_4 \kappa(A)$ where ϵ_4 can be chosen so that $r < 1$. We will also specify that ϵ_3 is chosen so that $\|\Delta A\| \leq \epsilon_4 \|A\|$ where $\Delta A = \widetilde{A} - A$ and $\widetilde{A}$ is the design matrix with respect to the $\tilde{y}_{i',t}$ series. Theorem 2.7.2 in Golub and Van Loan [1996] states that

$$\|\tilde{\beta} - \hat{\beta}\| \leq \frac{2\epsilon_4}{1 - r} \kappa(A) \|\hat{\beta}\|. \tag{10}$$

We can bound $\kappa(A)$ and $\hat{\beta}$ on E which allows one to choose ϵ_5 so that (10) is $\|\tilde{\beta} - \hat{\beta}\| < \epsilon_5$.

We now show that we can bound $|\hat{y}_{1,t}^{1,h} - \hat{\tilde{y}}_{i',t}^{1,h}|$ for $t + h > T^*$. We restrict attention to the case where $t + h > T^*$ because when $t + h \leq T^*$ the two competing forecasts $\hat{y}_{i,t}^{1,h}$ and $\hat{y}_{i,t}^{2,h}$ are the same in our setting for all i. Let $t + h > T^*$, then, when $t > T^*$,

$$\begin{aligned}
|\hat{y}_{1,t}^{1,h} - \hat{\tilde{y}}_{i',t}^{1,h}| &= |(\hat{\eta}_1 - \hat{\eta}_{i'}) + (\hat{\alpha}_1 - \hat{\alpha}_{i'}) + (\hat{\tilde{\phi}}_1 \hat{y}_{1,t-1}^{1,h-1} - \hat{\tilde{\phi}}_{i'} \hat{y}_{i',t-1}^{1,h-1}) + (\hat{\tilde{\theta}}_1' \mathbf{x}_{1,t} - \hat{\tilde{\theta}}_{i'}' \mathbf{x}_{i',t})| \\
&\leq 2\epsilon_5 + |\hat{\tilde{\theta}}_1' \mathbf{x}_{1,t} - \hat{\tilde{\theta}}_{i'}' \mathbf{x}_{i',t}| + |\hat{\tilde{\phi}}_1 \hat{y}_{1,t-1}^{1,h-1} - \hat{\tilde{\phi}}_{i'} \hat{y}_{i',t-1}^{1,h-1}| \\
&\leq 2\epsilon_5 + \|\mathbf{x}_{1,t}\| \|\hat{\tilde{\theta}}_1 - \hat{\tilde{\theta}}_{i'}\| + \|\hat{\tilde{\theta}}_{i'}\| \|\mathbf{x}_{1,t} - \mathbf{x}_{i',t}\| + \|\hat{y}_{1,t-1}^{1,h-1}\| |\hat{\tilde{\phi}}_1 - \hat{\tilde{\phi}}_{i'}\| + |\hat{\tilde{\phi}}_{i'}| |\hat{y}_{1,t-1}^{1,h-1} - \hat{y}_{i',t-1}^{1,h-1}| \\
&\leq 2\epsilon_5 + M\epsilon_5 + \epsilon_3 + \|\hat{y}_{1,t-1}^{1,h-1}\| \epsilon_5 + |\hat{\tilde{\phi}}_{i'}| |\hat{y}_{1,t-1}^{1,h-1} - \hat{y}_{i',t-1}^{1,h-1}|.
\end{aligned}$$

On E we can bound $|\hat{y}^{1,h-1}_{1,t-1}|$ for all $h = 1,\ldots,H$ and $t = 1,\ldots,T$, call this bound M_1. Let $\epsilon_6 = (2\epsilon_5 + M\epsilon_5 + \epsilon_3 + M_1\epsilon_5)$. Then,

$$\begin{aligned}
|\hat{y}^{1,h}_{1,t} - \hat{\tilde{y}}^{1,h}_{i',t}| &\leq 2\epsilon_5 + M\epsilon_5 + \epsilon_3 + |\hat{y}^{1,h-1}_{1,t-1}|\epsilon_5 + |\hat{\tilde{\phi}}_{i'}||\hat{y}^{1,h-1}_{1,t-1} - \hat{y}^{1,h-1}_{i',t-1}| \\
&\leq (2\epsilon_5 + M\epsilon_5 + \epsilon_3 + M_1\epsilon_5) + |\hat{\tilde{\phi}}_{i'}||\hat{y}^{1,h-1}_{1,t-1} - \hat{y}^{1,h-1}_{i',t-1}| \\
&\leq \epsilon_6 + |\hat{\tilde{\phi}}_{i'}||\hat{y}^{1,h-1}_{1,t-1} - \hat{y}^{1,h-1}_{i',t-1}| \\
&\leq \epsilon_6 + |\hat{\tilde{\phi}}_{i'}|\left(\epsilon_6 + |\hat{\tilde{\phi}}_{i'}||\hat{y}^{1,h-2}_{1,t-2} - \hat{y}^{1,h-2}_{i',t-2}|\right) \\
&\leq \epsilon_6(1 + |\hat{\tilde{\phi}}_{i'}|) + |\hat{\tilde{\phi}}_{i'}|^2|\hat{y}^{1,h-2}_{1,t-2} - \hat{y}^{1,h-2}_{i',t-2}| \\
&\leq \epsilon_6 \sum_{j=0}^{h} |\hat{\tilde{\phi}}_{i'}|^j + |\hat{\tilde{\phi}}_{i'}|^h|\hat{y}_{1,t-h} - \hat{y}_{i',t-h}| \\
&\leq \epsilon_6 \sum_{j=0}^{h} |\hat{\tilde{\phi}}_{i'}|^j + |\hat{\tilde{\phi}}_{i'}|^h\epsilon_3.
\end{aligned}$$

Let $\epsilon_7 = \max_{h=1,\ldots,H}\left(\epsilon_6 \sum_{j=0}^{h} |\hat{\tilde{\phi}}_{i'}|^j + |\hat{\tilde{\phi}}_{i'}|^h\epsilon_3\right)$ and conclude that

$$|\hat{y}^{1,h}_{1,t} - \hat{\tilde{y}}^{1,h}_{i',t}| < \epsilon_7, \tag{11}$$

for all $t = 1,\ldots,T$ and all $h = 1,\ldots,H$ such that $t + h > T^*$ on the event E. A similar argument with similar bounding choices to those which yielded (11) can be made when $t + h > T^*$ and $t \leq T^*$. A similar argument with similar bounding choices to those which yielded (11) can be made so that $|\hat{y}^{2,h}_{1,t} - \hat{\tilde{y}}^{2,h}_{i',t}| < \epsilon_7$.

We now bound the difference in loss differential vectors $\|\mathbf{d}_{1,t} - \mathbf{d}_{i',t}\|$ for $t > T^*$. What we have showed so far in (9) and (11) allows us to write

$$|y_{1,t} - \hat{y}^{1,h}_{1,t}| - |y_{1,t} - \hat{y}^{2,h}_{1,t}| \leq |\tilde{y}_{i',t} - \hat{\tilde{y}}^{1,h}_{i',t}| - |\tilde{y}_{i',t} - \hat{\tilde{y}}^{2,h}_{i',t}| + 2(\epsilon_3 + \epsilon_7). \tag{12}$$

We conclude that $\|\mathbf{d}_{1,t} - \mathbf{d}_{i',t}\| \leq \epsilon_8$ on E for all $t > T^*$ where ϵ_8 is chosen so that (12) holds when $\|\mathbf{d}_{1,t} - \mathbf{d}_{i',t}\| \leq \epsilon_8$ on E for all $t > T^*$. Putting this all together, we can pick ϵ_8 small enough so that for some fixed $\epsilon > 0$,

$$\left|\sqrt{T}\frac{\mathbf{w}_1'\bar{\mathbf{d}}^b_1 - \mathbf{w}_1'\bar{\mathbf{d}}_1}{\hat{\zeta}^b_1} - \sqrt{T}\frac{\mathbf{w}_{i'}'\bar{\mathbf{d}}^b_{i'} - \mathbf{w}_{i'}'\bar{\mathbf{d}}_{i'}}{\hat{\zeta}^b_{i'}}\right| < \epsilon \tag{13}$$

on the set E where we specify that $\mathbf{w}_1 = \mathbf{w}_{i'}$ and that the MBB procedures samples the same indices of blocks from $\mathbf{d}_{1,t}$ and $\mathbf{d}_{i',t}$, $t = T^* + 1,\ldots,T$ when forming the test statistics in (13). Thus for any fixed small $\epsilon < 0$ there exists a choice of σ^2 and ϵ_1 such that inferences involving the i' series for which post-shock behavior is observed can be substituted in place of the $i = 1$ series under study with high probability of a correspondence in inferences.

4 Numerical Examples

In this section, we illustrate shock prevalence testing using two real data examples. Section 4.1 discusses the prevalence testing of the shock that Conoco Phillips stock experienced on March 9th, 2020. Section 4.2 discusses the prevalence testing of the shock that seasonally adjusted log nonfarm payrolls experienced in April, 2020.

We will consider two models to analyze these datasets. The first model $\mathcal{M}_0$ is a simplification of the model (1) involving no changes to the post-shock dynamics. This model is explicitly written as,

$$y_{i,t} = \eta_i + \sum_{j=1}^{q_1} \phi_{i,j}y_{i,t-j} + \sum_{j=0}^{q_2-1} \boldsymbol{x}_{i,t-j}\boldsymbol{\theta}_{i,j+1} + \alpha_i D_{i,t} + \varepsilon_{i,t},$$

$$\alpha_i = \mu_\alpha + \sum_{j=0}^{q_2-1} \boldsymbol{x}_{i,T_i^*+1-j}\boldsymbol{\delta}_{i,j+1} + \varepsilon_{\alpha,i},$$

where $D_{i,t} = I(t > T_i^* + 1)$, $i = 1, \ldots, n+1$ and $t = 1, \ldots, T_i$. Model $\mathcal{M}_0$ describes the setting where one additive shock effect is permanent (persistent) over time, and we will refer to it as the "persistent" model. The second model under consideration is the "dynamic" model (1) which is denoted $\mathcal{M}$. Both models will specify that $q_1 = q_2$.

4.1 Testing shock prevalence for Conoco Phillips stock price series

As illustrated in Section 5.1 in Lin and Eck [2021], it is found that Conoco Phillips stock price experienced unprecedented shocks due to a Russia-OPEC oil supply control battle and increasing COVID-19 pandemic fear which were simultaneously realized on Monday, March 9th, 2020. In this data example, we are interested in investigating whether the shock studied in Lin and Eck [2021] is transient or prevalent. Denote the stock price of Conoco Phillips at time t as $y_{i,t}$ for ith time series, where $i = 1, \ldots n+1$, the time series of interest is ordered as the 1st time series, and n is the donor pool size. In the model of $y_{i,t}$, the covariates we consider are S&P 500 index prices, West Texas Intermediate (WTI) crude oil prices, dollar index, 13-week treasury bill rates, and Chicago Board Options Exchange volatility index (VIX).

We considered the shock effects that occurred on March 15, 2008, September 2008, and November 27, 2014, where the rationale behind this selection is discussed in Section 5.1(3) in Lin and Eck [2021]. However, we analyzed the multiple shock effects on September (September 8th, 2008; September 16th, 2008; and September 26, 2008) differently. Under the setting of our methodology, it is possible that the occurrence of those shocks can be modeled by one persistent shock effect or explained by that the model of $y_{i,t}$ changes after September 5th, 2008. In other words, we analyzed the three shock effects within one time series and considered three donors in our donor pool.

In this data example we set $T_i - T_i^* = 50$ and $K = 30$. Note that $K = 30$ is a sensible training sample size, see discussions for example in Lin and Eck [2021]. The quantity $T_i - T_i^* - \lceil 1.5 \cdot \sqrt{T_i - T_i^*} \rceil$ corresponds to the post-shock time length that analysts are interested in predicting whether the impact of the shock persists. We set $T_i - T_i^* = 50$ in order to make the post-shock time length moderately large and also for illustration purpose. For sensitivity analysis, we implemented the analysis for different $H = 2, 4, 6, 8$.

We summarize our numerical results in Table 1. In the Table 1, p_1 stands for the forecast comparison test of Quaedvlieg [2021] applied for adjusted and unadjusted forecasts on the time series of interest. Note that in practice, it cannot be computed because $y_{1,t}$ for $t > T_1^*$ is not observed. It is provided for retrospective analysis. Second, $\hat{p}_1$ is the estimate for p_1 using $\mathbf{W}^*$. We recorded the voting and weighted voting results in the form of indicators with one being voting for persistence and zero for transience. See the methodological details for $\hat{p}_1$, voting, and weighted voting in Section 3.2. In companion with the voting, the correctness of the voting is evaluated and provided in the brackets. For example, "0 (Yes)" implies that the procedure votes for transience and this decision is correct whereas "1 (No)" votes for persistence and it is incorrect.

Table 1: Results for Conoco Philips ($n = 3$), $\mathbf{W}^* = (0.221, 0, 0.779)$ from scaled covariates

H	Model	p_1	$\hat{p}_1$	$\lvert\hat{p}_1 - p_1\rvert$	Vote (Correct?)	weighted Vote (Correct?)
2	$\mathcal{M}_0$	1	0.023	0.977	0 (Yes)	1 (No)
	$\mathcal{M}$	1	1	0	0 (Yes)	0 (Yes)
4	$\mathcal{M}_0$	1	0.125	0.875	0 (Yes)	1 (No)
	$\mathcal{M}$	1	1	0	0 (Yes)	0 (Yes)
6	$\mathcal{M}_0$	0	0.999	0.999	0 (No)	0 (No)
	$\mathcal{M}$	1	1	0	0 (Yes)	0 (Yes)
8	$\mathcal{M}_0$	0	1	1	0 (No)	0 (No)
	$\mathcal{M}$	1	1	0	0 (Yes)	0 (Yes)

We now discuss the numerical results. Note that the synthetic weight vector $\mathbf{W}^*$ for the three donors is $(0.221, 0, 0.779)$. It suggests that the donor associated with the shock on November 27, 2014 is the most similar time series relative to the time series of interest, and the one on March 15, 2008 is the second. As pointed out by Lin and Eck [2021] in Section 5.1(3), the shocks experienced by the first and third donor are related to recessions, and oil supply shock, respectively. The results are intuitive in the sense that from a retrospective perspective of a data analyst, the shock experienced by the series $\{y_{1,t} : t = 1, \ldots, T_1\}$ is due to oil supply shock and recessions caused by COVID-19 pandemic.

Let a_i^1 and a_i^2 denote the AICs of $\mathcal{M}_0$ and $\mathcal{M}$ for ith time series, respectively. Based on (1), which describes $\mathcal{M}$ and the description of $\mathcal{M}$ in the beginning of Section 4, it may be inferred that $\mathcal{M}$ captures the post-shock dynamic change whereas $\mathcal{M}_0$ suggests a persistent post-shock additive location shift. Using the synthetic pairwise model selection procedures specified in Corollary 1, it turns out that $\sum_{i=2}^{4} w_i^* I(a_i^1 > a_i^2) = 0.779 > 0.5$. Using this aggregated model selection approach, we prefer $\mathcal{M}$ over $\mathcal{M}_0$. For reference, the AICs for $\mathcal{M}_0$ in the donors are $(348.261, 504.887, 377.073)$ whereas those for $\mathcal{M}$ in the donors are $(349.784, 505.012, 364.108)$.

From Table 1, we can see that p_1 for $\mathcal{M}$ is not sensitive to the changes of H and is uniformly 1. Under any reasonable significance level, based on the features of $\mathcal{M}$, we infer that post-shock dynamics do not differ from the pre-shock one and the shock is not persistent. In contrast, p_1 of $\mathcal{M}_0$ equals 1 for $H = 2, 4$ whereas equals 0 for $H = 6, 8$. It is possible that $\mathcal{M}_0$ does not fit the data well. In practice, a simple additive post-shock location-shift is rare and more complicated structures cannot be explained using this.

Two dashed vertical lines are plotted in Figure 1, where the left one corresponds to $t = T_1^*$ and the right one corresponds to $t = T_1^* + 1$. As shown by Lin and Eck [2021], $y_{1,t}$ experienced a substantial shock at $t = T_1^* + 1$ and it is evident from $y_{1,T_1^*} - y_{1,T_1^*+1}$ shown in Figure 1. We also graphed a benchmark horizontal dashed line for comparison, which corresponds to the stock price y_{1,T_1^*}. However, notice that even if $y_{1,t}$ first experienced a considerable drop at $t = T_1^* + 1$, it started to revert back and became closer to the benchmark horizontal line as t increases for $t > T_1^*$. Note that this observation confirms our model selection and what $\mathcal{M}$ suggests for shock transience. It is also important to re-emphasize that the test compares the forecasts from $t = T_1^* + H + \lceil 1.5 \cdot \sqrt{T_1 - T_1^*} \rceil$ to $t = T_1$. In other words, the beginning dramatic drop as seen from Figure 1 around March is not considered in the forecast comparison. It is reasonable to our objective as we are interested in long-term persistence rather-than short-term substantial changes.

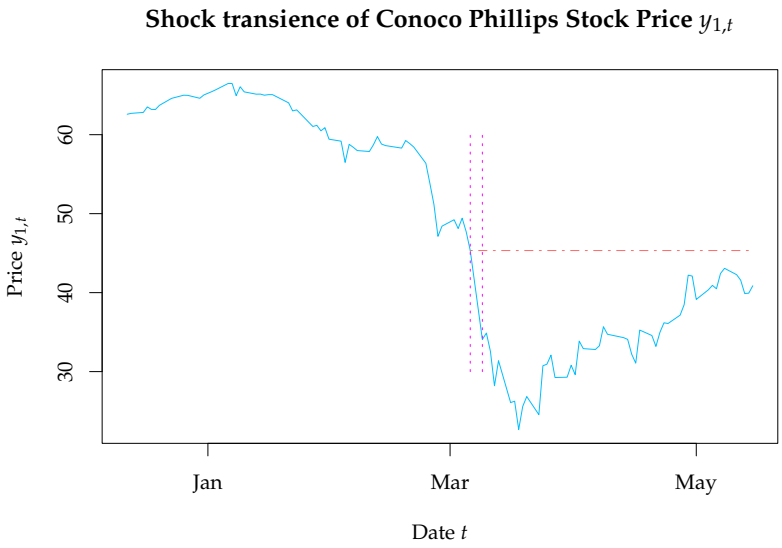

Figure 1: The shock transience of $\{y_{1,t} : 1, \ldots, T_1\}$

Next, we discuss the performance of $\hat{p}_1$, voting and weighted voting. From Table 1, we can observe that $|\hat{p}_1 - p_1|$ equals zero uniformly for $\mathcal{M}$ and $H = 2, 4, 6, 8$. It suggests that $\hat{p}_1$ under $\mathcal{M}$ serves as a favorable proxy for p_1 under $\mathcal{M}$. In contrast, $|\hat{p}_1 - p_1|$ is quite large for $\mathcal{M}_0$ and different H. It is possible that $\mathcal{M}_0$ does not fit the data properly or the donors are not representative of the time series of interest in the context of $\mathcal{M}_0$. Here, it is important to note that the comparability between donors and the time series of interest is model-dependent.

According to the last two columns of Table 1, the decisions yielded by voting and weighted voting are correct for $\mathcal{M}$ and $H = 2, 4, 6, 8$. However, the decisions for voting and weighted voting are incorrect for $\mathcal{M}_0$ and $H = 2, 4, 6, 8$ except for voting for $\mathcal{M}_0$ at $H = 2, 4$. Here, it is worth noting that weighted voting and voting may differ due to $\mathbf{W}^*$. Such difference may be valuable if the weight is concentrated in one donor that does not tell a consistent story to that of time series of interest. Next, the interpretations for these results are similar to those when we analyze $|\hat{p}_1 - p_1|$. Note that the decisions using $\hat{p}_1$ and weighted voting are similar. It is reasonable as they are functions of $\mathbf{W}^*$. Because p_1 is uniformly 1 for $\mathcal{M}$, it is reasonable to expect $\mathbb{P}(p_1 \leq .05)$ be small. Note that Proposition 1 and Corollary 1 show that the decisions made by voting and weighted voting procedures should be correct with very high probability if we assume the asymptotic approximation of the expected absolute error works well for $n = 3$.

4.1.1 Dual shock prevalence testing

We now introduce a forecast scenarios [Baumeister and Kilian, 2014b] approach to our prevalent testing framework. In Lin and Eck [2021] it was mentioned that Conoco Phillips possibly underwent a dual shock, which retrospectively, seems likely to be the case [Sharif et al., 2020]. There were also rumblings of a dual shock in real time [FitchRatings, 2020, Stevens, 2020, Westbrook, 2020]. In Lin and Eck [2021] a dual shock formulation was provided in which supposed financial shocks and oil supply shocks were separated to form their own donor pools, the additive shock methodology was applied to each of these donor pools, and the results were added together. This approach did very well, almost perfectly recovering the yet-to-be post-shock response.

We motivate a similar approach within our prevalence testing framework. The idea is to again separate out the financial and oil supply time series and add these series together which will create a new donor pool consisting of series which incorporate a dual shock whose behavior can be assessed with our methodology.

As in Lin and Eck [2021] the oil supply shock donor pool consists of the single November 27, 2014 series; and the financial shock donor pool consists of the remaining two series. These added series which account for a potential dual shock take the following form where the subscript o refers to the single November 27, 2014, the subscript j indexes the other series, and $T_o = T_j$ and $T_o^* = T_j^*$

$$\begin{aligned} z_{j,t} &= y_{o,t} + y_{j,t} \\ &= \bar{\eta}_j + (\phi z_{j,t-1} + \theta'_{o,t-1}\mathbf{x}_{o,t-1} + \theta'_{j,t-1}\mathbf{x}_{j,t-1})(1 - \bar{D}_t) \\ &\quad + (\bar{\alpha}_j + \tilde{\phi} z_{j,t-1} + \tilde{\theta}'_{o,t-1}\mathbf{x}_{o,t-1} + \tilde{\theta}'_{j,t-1}\mathbf{x}_{j,t-1})\bar{D}_t + \bar{\varepsilon}_t, \end{aligned} \tag{14}$$

where $\bar{\eta}_j = \eta_o + \eta_j$, $\bar{\alpha}_j = \alpha_o + \alpha_j$, $D_t = D_{0,t} = D_{j,t}$, and $\bar{\varepsilon}_t = \varepsilon_{0,t} + \varepsilon_{j,t} + D_t(\varepsilon_{\alpha,o} + \varepsilon_{\alpha,j})$. The model (14) is then a regression model with a heterogeneous error structure, the error distribution post-shock differs from the error distribution pre-shock. All model parameters are estimable in (14). Shock prevalence testing for $z_{j,t}$ is still conducted by comparing forecasts between adjusted and unadjusted conditional mean forecasts as illustrated in Section 3.2. However, aggregation strategies are slightly different for this model since covariates are included for two separate series.

We now examine distance-based weighting for model (14). Consider a weight vector $\mathbf{W}^d \in \mathbb{R}^{n-1}$ with elements $w_i \in [0,1]$ that satisfy $\sum_i w_i = 1$. Construct $\mathbf{X}_1^d = (\mathbf{x}'_{1,T_1^*+1},\ \mathbf{x}'_{1,T_1^*+1})$, $\mathbf{X}^d \in \mathbb{R}^{n-1\times 2p}$ with rows $(\mathbf{x}'_{o,T_o^*+1},\ \mathbf{x}'_{j,T_j^*+1})$, and $\hat{\mathbf{X}}_1^d(\mathbf{W}) = \mathbf{W}^{d\prime}\mathbf{X}^d$. Then the optimal weights satisfy

$$\mathbf{W}^{d*} = \arg\min_{\mathbf{W}\in\mathcal{W}} \|\mathbf{X}_1^d - \hat{\mathbf{X}}_1^d(\mathbf{W})\|_{2p}$$

where $\mathcal{W}$ is the set of all allowable weight vectors. Note that if the model is correctly specified, inference under this dual shock setup should work well if $\{y_{o,t} \colon t = 1, \ldots, T_0\}$ and $\{y_{j,t} \colon j = 2, \ldots, n, t = 1, \ldots, T_j\}$ are independent. We assume this is approximately satisfied because these time series are sufficiently separated out in time as introduced in the beginning of this section. We provide corresponding numerical results in Table 2.

Under the dual shock setup, the donor pool size of this dataset is $n = 2$. The fitted synthetic weight vector $\mathbf{W}^*$ turns out to be $(0.756, 0.244)$. It implies that the March 15, 2008 time series is more similar to the time series of interest than the September 2008 time series. Similarly, we conducted sensitivity analysis using $H = 2, 4, 6, 8$; and the results appear insensitive to H from Table 2. Based on Table 2, $\hat{p}_1$ is uniformly 1 so that we fail to detect shock persistence. The results of voting and weighted voting are consistent to this with both voting for shock transience.

Table 2: Results for Conoco Philips dual shock ($n = 2$), $\mathbf{W}^* = (0.756, 0.244)$ from scaled covariates

H	$\hat{p}_1$	Vote	weighted Vote
2	1	0	0
4	1	0	0
6	1	0	0
8	1	0	0

Unlike the results presented Table 1, we cannot know the underlying p_1 and evaluate whether the voting and weighted voting are correct. This is due to the fact that the time series of interest experience financial and oil shocks at the same time. As a result, the model is unidentifiable given the available data. Nevertheless, the shock transience evident in Figure 1 provides a reference and supports our claim. This result confirms what we obtained in Section 4.1 and sucessfully generalizes the dual shock in post-shock prediction [Lin and Eck, 2021] to prevalence testing.

4.2 Testing shock prevalence for Month-on-Month Change in Log Nonfarm Payrolls (Seasonally Adjusted)

What I'm trying to hit here

1. Why did I choose the donors I chose?
2. Why did I use the columns I did in OLS?
3. Why did I use the columns I did in the estimation of the points on the simplex?
4. Discuss the two models, M1, M2
5. Weighted AIC just as in previous example?
6. By weighted AIC, M1 is the preferred model

Arguably the most watched macroeconomic indicator for the US economy is the monthly release of the total nonfarm payrolls, which reflects the gains or losses of persons employed. However, its reputation as a lagging indicator of macroeconomic health constrains its usefulness. This makes it ripe for forecasting under the method herein. To this end, we examine the March 2020 lockdown measures as a shock to the US economy and labor market, enlisting previous macroeconomic shocks to the US economy as donors. The donor pool assembled notably contains no cases of lockdowns or even pandemics or epidemics. Instead, the common thread running through the time series under study and the donors is the element of a sudden and considerable negative surprise visited upon the economy from a source previously judged to be marginal.

We assemble a donor pool of substantial shocks to the US economy in the last four decades: the recessions of 1981-82, 2001-2003, and 2007-2009. These events bear similarity to the broad, seismic shock dealt to the US economy in March 2020 – and the labor market in particular – not only due to their unprecedented nature but also because of their potential to cause deep and lasting damage. The recession of 1981-82 resulted from Fed Chairman Volcker's unprecedented and controversial bid to halt persistently high inflation that had reigned for more than decade in the US [Sablik, 2013]. The economic shock that followed the September 11th terrorist both damaged consumer confidence and accelerated a contraction already in progress owing to the dot-com bubble [Kliesen, 2003]. Third and finally, the spring 2008 collapse of the investment bank Bear Stearns dealt a serious blow to confidence in the US financial system. Prior to that, financial markets had withstood the unprecedented declines in major US metro real estate markets that followed the peak of the market in 2006.

We used several monthly macroeconomic indicators as covariates for the linear model to arrive at an optimal weighting of the donor pool. These indicators are all transformed to their month-on-month change in log values, so as to be consistent with the time series under study and to capture the signal inherent in changes, not raw levels, of macroeconomic variables. These indicators are the US unemployment insurance transfers, real personal income, personal consumption expenditures, industrial production, consumer price index, Federal Reserve funds rate, and the count of black Americans age 20 and over on US nonfarm payrolls.

The variable selection steps are outlined as follows. At first, we have the same set of covariates $\{\mathcal{X}_1, \ldots, \mathcal{X}_p\}$ for each donor, where $\mathcal{X}_1$ corresponds to the name of the covariate. Next, for each donor, we conduct stepwise variable selection based on AIC so that for each donor, we end up with a model with covariates that may differ across donors, say $\{\mathcal{X}_j : j \in \mathcal{I}_i\}$ for $i = 2, \ldots, n+1$, where $\mathcal{I}_j$ is the collection of the indices corresponding to which covariate is selected. Next, we take the union $\cup_{i=2}^{n+1}\{\mathcal{X}_j : j \in \mathcal{I}_i\}$ to obtain the final set of covariates that are present in the model of every donor and the time series of interest.

In Table 3, using the synthetic pairwise model selection procedures specified in Corollary 1, because $\sum_{i=2}^{4} w_i^* I(a_i^1 > a_i^2) = 0$, we conclude that $\mathcal{M}_0$ is preferred. Thus, we turn our interest to that model.

Since p_1 is far above the rejection threshold for horizons H=1,2,3,4, we lack the evidence to reject the null of a transient shock in the time series under study at any horizon. The weighted p-value, $\hat{p}$, misses the ground truth p-value by no more than a modest amount at almost all horizons, and at some horizons is impressively close to the ground truth. Last, it is worth remarking the for the preferred model $\mathcal{M}_0$, we observe a monotone increasing sequence of ground-truth p-values as H increases. Such a phenomenon can be taken as evidence that as more data is harnessed, the decision is stable. The estimated p-value $\hat{p}_1$ do an adequate job at tracking this phenomenon as H increases. The performance of method is in spite of the unusual path that the pre-lockdown US economy took, where weeks of restrictions on human assocation and movement gave way to relaxation of controls and an economic boom.

Table 3: Results for Month-on-Month Change in Log Nonfarm Payrolls for $\mathcal{M}_0$ (Seasonally Adjusted) $(n = 3)$, $\mathbf{W}^* = (0.196, 0.804, 0.000)$ from scaled covariates

H	p_1	$\hat{p}_1$	$\lvert\hat{p}_1 - p_1\rvert$	Vote (Correct?)	weighted Vote (Correct?)
2	0.33	0.21	0.12	0 (Yes)	0 (Yes)
3	0.11	0.51	0.40	0 (Yes)	0 (Yes)
4	0.43	0.52	0.09	0 (Yes)	0 (Yes)
5	0.02	0.36	0.34	0 (No)	0 (No)
6	0.04	0.62	0.58	0 (No)	0 (No)
7	0.03	0.40	0.37	0 (No)	0 (No)
8	0.27	0.46	0.19	0 (Yes)	0 (Yes)
9	0.49	0.27	0.22	0 (Yes)	0 (Yes)
10	0.76	0.67	0.10	0 (Yes)	0 (Yes)

$\sum_{i=2}^{4} w_i^* I(a_i^1 > a_i^2) = 0.$
AICs for $\mathcal{M}_0$ in the donor pool are $(-609.173, -657.146, -678.949)$
AICs for $\mathcal{M}$ in the donor pool are $(-601.743, -649.484, -684.147)$

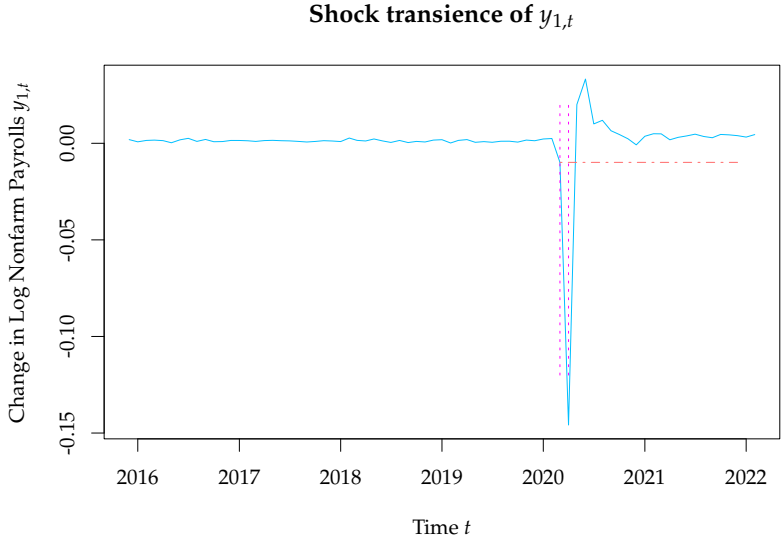

Figure 2: The shock transience of $\{y_{1,t} : 1, \ldots, T_1\}$

The pandemic-induced recession of 2020 coincided with a one-month drop in US nonfarm payrolls in excess of 20 million people, or 13.57%, between March and April 2020. This decline exceeds all other monthly declines by orders of magnitude. The distance-based weighting in [Lin and Eck, 2021], employing a convex combination of donor shocks, is known to be unable to extrapolate and thus recover the full shock effect. However, the procedure developed herein does not rely as heavily upon the full or partial recovery of the shock effect. What is potentially as important or more important is that the donor pool serve as an information source about the persistence of a shock effect in the time series under study, an information source based on the covariates that are posited to underlie the post-shock model dynamics. It is therefore the latter that may best explain the performance of the decision-making method developed here.

From Table 3, p_1 is insignificant for $H = 2, 3, 4$ and $H = 8, 9, 10$ but is significant for $H = 5, 6, 7$. Notice that weighted voting, vote, and $\hat{p}_1$ also does not perform well for $H = 5, 6, 7$. This implies something undetected and unique to the time series of interest occur. For this reason, we investigated the point estimates of the shock effects using data from $t = T_1^* + r - 37$ to $t = T_1^* + r$. Note that for different r, the training sample size is fixed and thus prevents training sample size to affect inference results. The results are presented in Table 4 with point estimates of the shock effects denoted by $\hat{\alpha}_{1,r}$ and corresponding standard error and p-value from t-test. Note that the t-test is significant for $r = 1, 2, 3, 4$ but insignificant for $r \geq 5$.

This naive test implies that the presence of a mean-shift shock effect predictably decays very quickly as more and more post-shock data is added for fitting purposes when, in truth, the shock is transient. However, the Quaedvlieg method runs $h = 1, ..., H$ ahead forecasts that will not necessarily indicate a quickly decaying shock due to the training data for the model including post-shock responses that have yet to decay. Thus the unadjusted and adjusted forecasts can deviate sharply when their underlying models are fit to a small amount of post-shock data. When H is small, there is a lot of post-shock data under consideration, and the shock decays quickly, the initial shock may not be detected by the Quaedvlieg method due to a loss-difference matrix evaluating several differences that are negligible far away from the initial shock. The same can be true when H is large. However, for medium size H, the influence of fitting the initial post-shock data may not yet accommodate the sharply decayed shock. The result of this is that the Quaedvlieg method can indicate a shifted significant difference between the adjusted and un-adjusted forecasts. (loss-difference matrix has dimensions $(T - T^* - \max(\lceil 1.5\sqrt{T - T^*} \rceil, H)) \times H$) When we have limited sample size, we should not consider large H.

Table 4: Point estimates of shock effects under $\mathcal{M}_0$ using data from $t = T_1^* + r - 37$ to $T_1^* + r$

r	$\hat{\alpha}_{1,r}$	$\sqrt{\hat{\mathbf{Var}}(\hat{\alpha}_{1,r})}$	p-value
1	−0.1344	0.0134	0.0000
2	−0.1379	0.0138	0.0000
3	−0.1921	0.0469	0.0002
4	−0.0055	0.0065	0.4056
5	0.0054	0.0041	0.1950
6	−0.0013	0.0025	0.6213
7	−0.0042	0.0025	0.1003
8	−0.0028	0.0022	0.2109
9	−0.0010	0.0020	0.6203
10	0.0005	0.0019	0.7788
11	0.0006	0.0018	0.7636
12	0.0029	0.0019	0.1368
13	0.0024	0.0022	0.2823
14	0.0025	0.0021	0.2442
15	0.0015	0.0021	0.4883
16	0.0008	0.0023	0.7245
17	0.0013	0.0022	0.5693
18	0.0015	0.0022	0.4907

4.3 Simulations

In this section, we propose several simulation setups to illustrate the performance of our methods and find when they work well and when do not. The parameter setup is shown as follows.

$$
\begin{aligned}
\phi_{i,j}, \theta_{i,j} &\overset{iid}{\sim} U(-0.4, 0.4) \\
\delta_{i,j} &\overset{iid}{\sim} N(1, 0.5^2) \\
\varepsilon_{\alpha,i} &\overset{iid}{\sim} N(0, 2) \\
T_i^* &\overset{iid}{\sim} \min\left(90, \left[Gamma(15, 10)\right]\right) \\
\varepsilon_{i,t} &\overset{iid}{\sim} N(0, \sigma) \\
X_{i,j} &\overset{iid}{\sim} Gamma(1, 2) \\
e_{i,j}, e_{i,j}^* &\overset{iid}{\sim} beta(\alpha_i^*, \beta_i^*) \\
\alpha_i^* &= a_1 \|\mathbf{x}_{i,T_i^*+1}\| + b_1 \\
\beta_i^* &= a_2 \|\mathbf{x}_{i,T_i^*+1}\| + b_2 \\
\tilde{\theta}_{i,j} &= \theta_{i,j} + \frac{1}{5} \cdot \left(e_{ij} - \frac{1}{2}\right) \\
\tilde{\phi}_{i,j} &= \phi_{i,j} + \frac{1}{5} \cdot \left(e_{ij}^* - \frac{1}{2}\right)
\end{aligned}
$$

Note that $\phi_{i,j}, \theta_{i,j} \overset{iid}{\sim} U(-0.4, 0.4)$ are designed to prevent non-stationarity. The reason for the design of T_i^* is similar to that in Section 4 of Lin and Eck [2021]. In particular, we focus on varying μ_α, σ, and a_1 under the design of $\mathcal{M}$ as specified in Section 3.1. Note that μ_α controls the signal of the shock whereas

σ manages the noise of the response. a_1, a_2 is related to the magnitude of α_i^*, which depends also on $\|\mathbf{x}_{i,T_i^*+1}\|$. This simulation design allows for the situations when the post-shock dynamics depends on the covariates at the time points the shock occur. For simplicity, we set $a_2 = 0$ and fix $b_1 = b_2 = 1$. By varying a_1, we know how our methods perform as $\mathbb{E}(e_{i,j})$ ranges from 0 to 1.

For illustration purpose, we keep other parameters simple and friendly for easier computation with bootstrap sample size $B = 200$, maximum horizon $H = 10$, autoregressive parameter order $q_1 = q_2 = 2$, donor pool sample size $n = 10$, and covariate dimension $p = 2$. Note that we set the block size in the moving block bootstrap to be $\ell = 4$ and the bandwidth for HAC estimator of the covariance matrix to be 4 (see Section 3.2).

In the first simulation, we vary μ_α and a_1 with results presented in Table 5. First, fixing a_1, we see that the $\hat{\mathbb{E}}(\hat{p})$ and the misclassification rate (MCR) of voting and weighted voting get improved as μ_α increases. When $\mu_\alpha = 0$, the misclassification rates are close to 0.5. It makes sense by Proposition 1 and 2 since $\hat{\mathbb{E}}(p_1)$ is close to 0.5 in this scenario. Under this design, the shock signal is neither strong enough or weak enough for the method to classify it as being persistent or transient so that the results may be poor. However, as μ_α increases, the shock signal strength gets large. Table 5 shows that the results appear satisfactory when $\mu_\alpha = 100$.

Second, fixing μ_α, varying a_1 lead to a better fitting results though the improvement is marginal and does not hold for some combinations of μ_α and a_1. Note that as a_1 increases from 10^{-3} to 10^3, $\mathbb{E}(e_{ij})$ tends from 0 to ∞. According to $\mathcal{M}$ (see Section 3.1), this change make the post-shock dynamics more different than the pre-shock dynamics so that the shock signal strength gets larger.

Table 5: Monte Carlo Simulation of $\mathcal{M}$ with varying μ_α and a_1

μ_α	a_1	$\hat{\mathbb{E}}(\hat{p})$	$\hat{\mathbb{E}}(p_1)$	$\hat{\mathbb{E}}(\lvert\hat{p} - p_1\rvert)$	MCR of voting	MCR of weighted voting
0		0.458 (0.025)	0.420 (0.033)	0.420 (0.024)	0.495 (0.035)	0.440 (0.035)
10	10^{-3}	0.206 (0.022)	0.195 (0.027)	0.278 (0.025)	0.250 (0.031)	0.345 (0.034)
100		0.037 (0.010)	0.010 (0.007)	0.047 (0.012)	0.010 (0.007)	0.035 (0.013)
0		0.382 (0.025)	0.343 (0.031)	0.423 (0.025)	0.470 (0.035)	0.520 (0.035)
10	1	0.185 (0.020)	0.141 (0.024)	0.267 (0.025)	0.175 (0.027)	0.285 (0.032)
100		0.014 (0.006)	0.025 (0.011)	0.039 (0.013)	0.025 (0.011)	0.035 (0.013)
0		0.365 (0.024)	0.268 (0.029)	0.401 (0.023)	0.430 (0.035)	0.450 (0.035)
10	10^3	0.130 (0.017)	0.139 (0.024)	0.235 (0.024)	0.160 (0.026)	0.250 (0.031)
100		0.022 (0.007)	0.010 (0.007)	0.032 (0.010)	0.010 (0.007)	0.030 (0.012)

Table 6: Monte Carlo Simulation with varying σ and a_1 ($B = 200$, $\mu_\alpha = 20$, $H = 10$, $\ell = 4$, bandwidth 4, $q_1 = q_2 = 2$, $b_1 = b_2 = 1$, $a_2 = 0$, $n = 10$, $p = 2$, $\sigma_\alpha = 2$, $\sigma_\delta = 0.05$, $\mu_\delta = 1$)

σ	a_1	$\hat{\mathbb{E}}(\hat{p})$	$\hat{\mathbb{E}}(p_1)$	$\hat{\mathbb{E}}\big(\lvert\hat{p} - p_1\rvert\big)$	MCR of voting	MCR of weighted voting
0.01		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
1		0.116 (0.016)	0.117 (0.022)	0.191 (0.023)	0.120 (0.023)	0.170 (0.027)
5	10^{-3}	0.545 (0.026)	0.581 (0.034)	0.471 (0.026)	0.440 (0.035)	0.475 (0.035)
10		0.685 (0.023)	0.713 (0.031)	0.431 (0.026)	0.330 (0.033)	0.415 (0.035)
50		0.935 (0.013)	0.935 (0.017)	0.119 (0.020)	0.065 (0.017)	0.095 (0.021)
0.01		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
1		0.112 (0.016)	0.090 (0.020)	0.192 (0.023)	0.105 (0.022)	0.195 (0.028)
5	1	0.493 (0.026)	0.424 (0.034)	0.499 (0.025)	0.460 (0.035)	0.520 (0.035)
10		0.626 (0.025)	0.610 (0.034)	0.430 (0.026)	0.400 (0.035)	0.380 (0.034)
50		0.961 (0.009)	0.891 (0.022)	0.136 (0.022)	0.105 (0.022)	0.125 (0.023)
0.01		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
1		0.078 (0.013)	0.085 (0.020)	0.149 (0.021)	0.090 (0.020)	0.145 (0.025)
5	10^3	0.461 (0.025)	0.481 (0.035)	0.530 (0.025)	0.480 (0.035)	0.570 (0.035)
10		0.684 (0.023)	0.546 (0.035)	0.473 (0.026)	0.430 (0.035)	0.480 (0.035)
50		0.921 (0.014)	0.894 (0.022)	0.172 (0.023)	0.105 (0.022)	0.155 (0.026)

In the second simulation, we fix μ_α to be 20 and try to vary σ and a_1. The results are presented in Table 6. Fixing a_1, we can observe that the results become worse and then better as σ increases. Notably, when $\sigma = 0.01$, which is almost close to zero as compared to $\mu_\alpha = 20$, $\hat{\mathbb{E}}\big(\lvert\hat{p} - p_1\rvert\big)$ and misclassification rates of voting and weighted voting are almost zero. This implies that when the donors are nearly replicates of the time series of interest, our method is almost perfect. When σ is small, the shock signal from μ_α is relatively large and can be easily captured as a persistent shock. In contrast, when σ is large, as μ_α is fixed, the shock signal is relatively small so that it is understood as a transient shock. In terms of the U shaped performance of our results, we posit that when σ is moderately large, the shock signal can be easily understood as persistent or transient because under this configuration, it is very possible that the response $y_{i,t}$ can attain a magnitude greater than the shock $\alpha_{i,t}$ or smaller.

In Table 6, fixing σ, we do not observe a clear trend about how a_1 interacts with $\hat{\mathbb{E}}\big(\lvert\hat{p} - p_1\rvert\big)$ and the misclassification rates of voting and weighted voting. Note that a_1 controls the mean and variance of the noise for the post-shock autoregressive parameters of $\tilde{\theta}_{i,j}, \tilde{\phi}_{i,j}$. It appears that a_1 does not affect the results, suggesting that our method is not sensitive to the noise of the unobserved parameter distribution.

In conclusion, via the simulation, we find that our method works well when the shock signal is either very strong or weak compared to the noise of the response. However, the signal may mislead our method when the shock signal strength is similar to that of the noise. Note that both $\tilde{\theta}_{i,j}, \tilde{\phi}_{i,j}$ and μ_α play a role in the post-shock dynamics but $\tilde{\theta}_{i,j}, \tilde{\phi}_{i,j}$ controls the structure. Based on our results, we find that our method is not sensitive to the change of structure but instead the mean shift due to μ_α.

5 Discussion

needs work Talk about post-shock forecasting as parameterizing exogeneity and using hindsight in the donor pool to explicitly estimate exogeneity apriori. This framing has its limits, we are not able to perform post-shock forecasts when we are temporarily removed from the known shock time. For example, in the Conoco Philips analysis it is hard to imagine a successful post-shock forecast being made in the days

prior to the weekend preceding Monday, March 9th 2020 as the OPEC-Russia oil supply increases and Covid-19 fears had not yet taken shape.

needs work The post-shock forecasting methodology presented here and in Lin and Eck [2021] is heavily dependent on the construction of a suitable donor pool which will exhibit a degree of subjectivity. This is a limitation of our approach, but we argue that this limitation is far from a deal breaker and has its virtues. Our approach allows for subject matter experts to obtain answers in a tractable manner from donor pool to model to answers. Subject matter experts can therefore investigate and debate the appropriateness and effect that particular donors have on the conclusions which are obtained from the method. This invites a robustness possibility in which the inclusion/exclusion of specific donors do not materially effect the decision that are obtained.

needs work One hand, Agarwal et al. [2020a] do not require such a parametric specification. On the other hand, the success of their approach is retrospective and its success is demonstrated in panel settings in which multiple donors are needed to inform hypothetical trajectories concerning the exact same shock.

6 Mathematical Appendix

Proof of Proposition 1:

Proof. Notice that as $\mathbb{P}(p_i \leq \alpha) = \kappa$ for $i = 1, \ldots, n+1$ and $I(p_i \leq \alpha)$ is a Bernoulli random variable, $I(p_i \leq \alpha)$ is identically distributed for $i = 1, \ldots, n+1$. Since p_i is pairwise independent and p_i are identically distributed for $i = 2, \ldots, n+1$, by Weak Law of Large Numbers,

$$\frac{1}{n}\sum_{i=2}^{n+1} I(p_i \leq \alpha) \xrightarrow{p} \mathbb{P}(p_i \leq \alpha) = \mathbb{P}(p_1 \leq \alpha),$$

Define

$$f\colon [0,1] \mapsto \{0,1\} \text{ with } f(x) = I(x \geq 0.5).$$

Let $C(f)$ denote the continuity set of f. Suppose that $\mathbb{P}(p_1 \leq \alpha) \neq 0.5$. In this case, notice that

$$\mathbb{P}(\mathbb{P}(p_1 \leq \alpha) \in C(f)) = 1.$$

By Slutsky's Theorem, we have

$$I\left\{\frac{1}{n}\sum_{i=2}^{n+1} I(p_i \leq \alpha) \geq 0.5\right\} \xrightarrow{p} I\{\mathbb{P}(p_1 \leq \alpha) \geq 0.5\}.$$

It follows that

$$I\left\{\frac{1}{n}\sum_{i=2}^{n+1} I(p_i \leq \alpha) \geq 0.5\right\} - I(p_1 \leq \alpha) \xrightarrow{p} I\{\mathbb{P}(p_1 \leq \alpha) \geq 0.5\} - I(p_1 \leq \alpha)$$

Since the function $g(x) = |x|$ is continuous in x, by continuous mapping theorem,

$$\left|I\left\{\frac{1}{n}\sum_{i=2}^{n+1} I(p_i \leq \alpha) \geq 0.5\right\} - I(p_1 \leq \alpha)\right| \xrightarrow{p} |I\{\mathbb{P}(p_1 \leq \alpha) \geq 0.5\} - I(p_1 \leq \alpha)|$$

Moreover, note that

$$\mathbb{E}\left\{|I\{\mathbb{P}(p_1 \leq \alpha) \geq 0.5\} - I(p_1 \leq \alpha)|\right\} = \begin{cases} 1 - \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) > 0.5 \\ \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) < 0.5 \end{cases}$$

$$\leq 0.5.$$

Due to the fact that

$$\left| I\left\{ \frac{1}{n}\sum_{i=2}^{n+1} I(p_i \leq \alpha) \geq 0.5 \right\} - I(p_1 \leq \alpha) \right| \leq 1$$

is bounded by 1, by Dominated Convergence Theorem for the version of convergence in measure,

$$\left| I\left\{ \frac{1}{n}\sum_{i=2}^{n+1} I(p_i \leq \alpha) \geq 0.5 \right\} - I(p_1 \leq \alpha) \right| \xrightarrow{\mathcal{L}_1} |I\{\mathbb{P}(p_1 \leq \alpha) \geq 0.5\} - I(p_1 \leq \alpha)|.$$

That would imply that

$$\mathbb{E}\left\{ \left| I\left\{ \frac{1}{n}\sum_{i=2}^{n+1} I(p_i \leq \alpha) \geq 0.5 \right\} - I(p_1 \leq \alpha) \right| \right\} \to \begin{cases} 1 - \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) > 0.5 \\ \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) < 0.5 \end{cases}$$

That is, the expected misclassification rate of voting converges to

$$\begin{cases} 1 - \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) > 0.5 \\ \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) < 0.5 \end{cases}$$

□

We now provide a proof for Proposition 2. This proof requires a lemma from probability theory whose proof we include for completeness.

Lemma 1. *For each $n \in \mathbb{N}$, let $c_{n1}, \ldots, c_{nn}$ be real numbers bounded by some $K > 0$, and let $X_{n1}, \ldots, X_{nn}$ be pairwise independent random variables defined on a probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, and let $\mathbb{E}_n$ and $\mathbf{Var}_n$ denote the corresponding expectation and variance. If (b_n) is a sequence of positive numbers such that $b_n \uparrow \infty$ such that*

$$\sum_{i=1}^{n} \mathbb{P}_n(|X_{ni}| > b_n) \to 0 \quad \text{and} \quad \frac{1}{(b_n)^2}\sum_{i=1}^{n} \mathbb{E}_n[X_{ni}^2; |X_{ni}| \leq b_n] \to 0,$$

then

$$\frac{1}{b_n}\sum_{i=1}^{n} c_{ni}(X_{ni} - \mathbb{E}_n[X_{ni}; X_{ni} \leq b_n]) \xrightarrow{p} 0.$$

Proof. For each $n \in \mathbb{N}$ set

$$S_n = \sum_{i=1}^{n} c_{ni}X_{ni}, \quad T_n = \sum_{i=1}^{n} c_{ni}Y_{ni}, \quad Y_{ni} = X_{ni}I(|X_{ni}| \leq b_n), \quad i = 1, \ldots, n.$$

The goal is to show that $(S_n - \mathbb{E}_n[T_n])/b_n \xrightarrow{p} 0$. For this, it suffices to show that (1) $(T_n - S_n)/b_n \xrightarrow{p} 0$ and (2) $(T_n - \mathbb{E}_n[T_n])/b_n \xrightarrow{p} 0$ because

$$(T_n - \mathbb{E}_n[T_n])/b_n - (T_n - S_n)/b_n = (S_n - \mathbb{E}_n[T_n])/b_n.$$

We first prove (1). Notice that for every $\varepsilon > 0$ and $n \in \mathbb{N}$,

$$\{|S_n - T_n| > \varepsilon\} \subseteq \bigcup_{i=1}^{n} \{X_{ni} \neq Y_{ni}\} = \bigcup_{i=1}^{n} \{|X_{ni}| > b_n\},$$

and, as a result, by Boole's inequality we have

$$\mathbb{P}_n(|S_n - T_n| > \varepsilon) \leq \sum_{i=1}^{n} \mathbb{P}_n(|X_{ni}| > b_n) \to 0,$$

which proves (1). Next, we prove (2). Since $\mathcal{L}^2$ implies convergence in probability, it suffices to show that $(T_n - \mathbb{E}[T_n])/b_n \xrightarrow{\mathcal{L}^2} 0$, i.e., $\mathbf{Var}[T_n]/(b_n)^2 \to 0$. Since $X_{n1}, \ldots, X_{nn}$ are pairwise independent, so are $Y_{n1}, \ldots, Y_{nn}$, and consequently

$$\mathrm{Var}_n[T_n] = \sum_{i=1}^{n} \mathrm{Var}_n[Y_{ni}] \leq \sum_{i=1}^{n} \mathbb{E}_n[(c_{ni}X_{ni})^2; |X_{ni}| \leq b_n] \leq K^2 \sum_{i=1}^{n} \mathbb{E}_n[(X_{ni})^2; |X_{ni}| \leq b_n].$$

As a result,

$$0 \leq \frac{\mathrm{Var}_n[T_n]}{b_n^2} \leq K^2 \cdot \frac{1}{b_n^2} \sum_{i=1}^{n} \mathbb{E}_n[(X_{ni})^2; |X_{ni}| \leq b_n] \to 0,$$

which proves the result by sandwich theorem. □

We now have enough to prove Proposition 2:

Proof. The proof is rather similar to Proposition 1. It suffices to show that

$$\sum_{i \in \mathcal{I}_n} w_i I(p_i \leq \alpha) \xrightarrow{p} \kappa_1 = \mathbb{P}(p_1 \leq \alpha).$$

and the remaining proof is the same as that of Proposition 1 in terms of applying Dominated Convergence Theorem in the version of convergence in probability. As p_i are pairwise independent for $i \in \mathcal{I}_n$, $I(p_i \leq \alpha)$ are pairwise independent for $i \in \mathcal{I}_n$. The idea is to prove the required condition of Lemma 1 holds. As $\mathcal{I}_n$ is non-empty, $|\mathcal{I}_n| \to \infty$, $b_n > 1$, and $b_n \to \infty$ as $n \to \infty$,

$$\sum_{i \in \mathcal{I}_n} \mathbb{P}\big(I(p_i \leq \alpha) > b_n\big) \to 0$$
$$\frac{1}{b_n^2} \sum_{i \in \mathcal{I}_n} \mathbb{E}\big[I(p_i \leq \alpha) \,|\, I(p_i \leq \alpha) \leq b_n\big] = \frac{1}{b_n^2} \sum_{i \in \mathcal{I}_n} \mathbb{P}\big(p_i \leq \alpha\big) = \frac{1}{b_n^2} \sum_{i \in \mathcal{I}_n} w_i \kappa_i \to 0$$

because $\sum_{i \in \mathcal{I}_n} w_i \kappa_i = O(1)$ by the assumption $\sum_{i \in \mathcal{I}_n} w_i \kappa_i \to \kappa_1$. Let $c_{ni} = w_i b_n$ for $i \in \mathcal{I}_n$. Since $w_i b_n \leq K$ for some $K > 0$ and b_n, we have

$$\begin{aligned} & \frac{1}{b_n} \sum_{i \in \mathcal{I}_n} c_{ni}\big(I(p_i \leq \alpha) - \mathbb{P}(I(p_i \leq \alpha) \,|\, I(p_i \leq \alpha) \leq b_n)\big) \\ = \; & \frac{1}{b_n} \sum_{i \in \mathcal{I}_n} c_{ni}\big(I(p_i \leq \alpha) - \mathbb{P}(I(p_i \leq \alpha)\big) \xrightarrow{p} 0 \end{aligned}$$

which follows from Lemma 1. The above is equivalent to

$$\sum_{i \in \mathcal{I}_n} w_i I(p_i \leq \alpha) - \sum_{i \in \mathcal{I}_n} w_i \kappa_i \xrightarrow{p} 0$$

As $\sum_{i \in \mathcal{I}_n} w_i \kappa_i \to \kappa_1$ as $n \to \infty$, by Slutsky's Theorem,

$$\sum_{i \in \mathcal{I}_n} w_i I(p_i \leq \alpha) \xrightarrow{p} \kappa_1,$$

which finishes the proof. □

References

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505, 2010.

Anish Agarwal, Abdullah Alomar, Arnab Sarker, Devavrat Shah, Dennis Shen, and Cindy Yang. Two burning questions on covid-19: Did shutting down the economy help? can we (partially) reopen the economy without risking the second wave? *arXiv preprint arXiv:2005.00072*, 2020a.

Anish Agarwal, Devavrat Shah, and Dennis Shen. Synthetic interventions. *arXiv preprint arXiv:2006.07691*, 2020b.

Badi H Baltagi. Forecasting with panel data. *Journal of forecasting*, 27(2):153–173, 2008.

Christiane Baumeister and Lutz Kilian. A general approach to recovering market expectations from futures prices with an application to crude oil. 2014a.

Christiane Baumeister and Lutz Kilian. Real-time analysis of oil price risks using forecast scenarios. *IMF Economic Review*, 62(1):119–145, 2014b.

Michael P Clements, Ana Beatriz Galvão, et al. Measuring the effects of expectations shocks. Technical report, Economic Modelling and Forecasting Group, 2019.

Dean Croushore and Charles L Evans. Data revisions and the identification of monetary policy shocks. *Journal of Monetary Economics*, 53(6):1135–1160, 2006.

Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 13(3):253–263, 1995.

FitchRatings. Oil shock compounds sovereign credit risks from coronavirus. https://www.fitchratings.com/research/sovereigns/oil-shock-compounds-sovereign-credit-risks-from-coronavirus-09-03-2020, 2020.

Raffaella Giacomini and Halbert White. Tests of conditional predictive ability. *Econometrica*, 74(6): 1545–1578, 2006.

Gene H Golub and Charles F Van Loan. *Matrix Computations (third edition)*. Johns Hopkins University Press, 1996.

Andre J Hoogstrate, Franz C Palm, and Gerard A Pfann. Pooling in dynamic panel-data models: An application to forecasting gdp growth rates. *Journal of Business & Economic Statistics*, 18(3):274–283, 2000.

Lutz Kilian and Helmut Lütkepohl. *Structural vector autoregressive analysis*. Cambridge University Press, 2017.

Kevin L Kliesen, 2003. URL https://files.stlouisfed.org/files/htdocs/publications/review/03/09/Kliesen.pdf.

Gary Koop and Dimitris Korobilis. Forecasting inflation using dynamic model averaging. *International Economic Review*, 53(3):867–886, 2012.

Hans R Künsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, pages 1217–1241, 1989.

Jilei Lin and Daniel J Eck. Minimizing post-shock forecasting error through aggregation of outside information. *International Journal of Forecasting*, 2021.

Laura Liu, Hyungsik Roger Moon, and Frank Schorfheide. Forecasting with dynamic panel data models. *Econometrica*, 88(1):171–201, 2020.

Regina Y Liu and Kesar Singh. Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the limits of bootstrap*, 225:248, 1992.

Francesca Monti. Forecast with judgment and models. *National Bank of Belgium Working Paper*, (153), 2008.

M Hashem Pesaran, Yongcheol Shin, and Ron P Smith. Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American statistical Association*, 94(446):621–634, 1999.

Rogier Quaedvlieg. Multi-horizon forecast comparison. *Journal of Business & Economic Statistics*, 39 (1):40–53, 2021.

Venkatram Ramaswamy, Wayne S DeSarbo, David J Reibstein, and William T Robinson. An empirical pooling approach for estimating marketing mix elasticities with pims data. *Marketing Science*, 12(1): 103–124, 1993.

Tim Sablik, Nov 2013. URL https://www.federalreservehistory.org/essays/recession-of-1981-82.

Arshian Sharif, Chaker Aloui, and Larisa Yarovaya. Covid-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the us economy: Fresh evidence from the wavelet-based approach. *International Review of Financial Analysis*, 70:101496, 2020.

Pippa Stevens. Oil plunges 24% for worst day since 1991, hits multi-year low after opec deal failure sparks price war. https://www.cnbc.com/2020/03/08/oil-plummets-30percent-as-opec-deal-failure-sparks-price-war-fears.html, 2020. Published on 2020-03-08 at 6:03 PM EDT; Accessed on 2022-06-03.

Lars EO Svensson. Monetary policy with judgment: Forecast targeting. Technical report, National Bureau of Economic Research, 2005.

Tom Westbrook. Forex-dollar and commodity currencies trampled by oil, virus double shock. https://www.reuters.com/article/global-forex-idUSL4N2B21OH, 2020. Published on 2020-03-09 at 12:45 AM; Accessed on 2022-06-03.