

# 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 motivated by the methodology of Lin and Eck [2021] and the multi-horizon forecast testing procedure of Quaadvlieg [2021] in order to test whether or not time series which incorporate shock information exhibit superior predictive performance over forecasts which do not over a set time horizon. This effectively tests whether a shock is transient or permanent over a set time horizon.

**This needs work** The core idea of our methodology is to sensibly aggregate similar past realized shock effects which arose from other time series, and then incorporate the aggregated shock effect estimator into the present forecast. Our testing method embraces ideas from forecast aggregation in the post-shock

setting [Lin and Eck, 2021], forecast comparison [Diebold and Mariano, 1995, Quaadvlieg, 2021], and p-value combination. It also has 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, Agarwal et al., 2020], expectation shocks [Croushore and Evans, 2006, Baumeister and Kilian, 2014a, Clements et al., 2019].

Need more pointed references on similar methods. Need to discuss how this method extends Lin and Eck [2021] and is an application of Quaadvlieg [2021].

## 2 Setting

We will suppose that a researcher has multivariate time series data  $\mathbf{y}_{i,t}$ ,  $t = 1, \dots, T_i$  and  $i = 1, \dots, n+1$ . 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 is 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  $T_i^* \leq T_i + H$ , and we will suppose that  $T_1^* = T_1 + 1$  and  $1 < T_i^* < T_i + H$  for  $i \geq 2$ . We will specify that  $\mathbf{x}_{i,t=T_i^*}$  is observed before the shock takes effect on  $y_{i,t=T_i^*}$ .

We are interested in point forecasts  $y_{i,t}^h$  at multiple horizons,  $h = 1, \dots, H$  with the aim of determining whether the shock has an effect on  $y_{i,t}^h$ . Quaadvlieg [2021] provided a methodology for comparing forecasts jointly across all horizons of a forecast path,  $h = 1, \dots, H$ . In our post-shock setting, we want to compare the forecasts

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

where  $\hat{y}_{i,t}^{1,h}$  is the forecast for  $y_{i,t}$  that accounts for the yet-to-be observed structural 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. We will compare these forecasts in terms of their loss differential

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

where  $\mathbf{L}_{i,t,j} \in \mathbb{R}^H$  has elements  $L^h(y_{i,t}, \hat{y}_{i,t}^{j,h})$ ,  $j = 1, 2$ , and  $L$  is a loss function. Hypothesis tests in Quaadvlieg [2021] are with respect to  $E(\mathbf{d}_{i,t}) = \mu_{i,t}$ . Conditions for these tests require conditions of Giacomini and White [2006].

We will be interested in  $\mu_i = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T \mu_{i,t}$ .

**Note:** We need more formality for constructing  $\hat{y}_{1,t}^{1,h}$ . We could use the forecasts in Lin and Eck [2021] and then consider  $h$ -ahead methods after adjusting for the shock. Or we could consider aggregation approaches which average all post-shock responses of the series in the donor pool.

We will consider the average superior predictive ability (aSPA) to assess whether or not a shock is permanent or transitory. 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.

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

$$H_{i,\text{aSPA}}^0 : \mu_i^{(\text{Avg})} \leq 0. \quad (1)$$

A studentized test statistic corresponding to the null hypothesis (1) is of the form

$$t_{i,\text{aSPA}} = \frac{\sqrt{T_i} \bar{d}_i}{\hat{\zeta}_i}, \quad (2)$$

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  $\Omega_i = \text{avar}(\sqrt{T_i}(\bar{d}_i - \mu_i))$ .

We now demonstrate that computing p-values corresponding to the hypothesis test (1) using the moving block bootstrap (MBB) is a good approximation. Let  $l$  be the block length of the MBB where we assume that  $T_i = lK_i$ . Let  $I_1, \dots, I_{K_i}$  be iid random variables uniformly distributed on  $\{1, \dots, T_i - l + 1\}$ , and define the array

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

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

$$\left(\hat{\omega}_i^{hb}\right)^2 = \frac{1}{K_i} \sum_{k=1}^{K_i} \left[ \frac{1}{l} \left( \sum_{t=1}^l d_{i,(k-1)l+t}^{hb} - \bar{d}_i^{hb} \right)^2 \right], \quad (3)$$

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

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}^b \left[ \sqrt{T_i} \frac{\mathbf{w}_i' \bar{\mathbf{d}}_i^b - \mathbf{w}_i' \bar{\mathbf{d}}_i}{\hat{\zeta}_i^b} \right] - \mathbb{P} \left[ \sqrt{T_i} \frac{\mathbf{w}_i' \bar{\mathbf{d}}_i - \mathbf{w}_i' \mu_i}{\hat{\zeta}_i^b} \right] \right| \xrightarrow{P} 0, \quad (4)$$

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

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

The results (4) implies that

$$|\hat{p}_i^B - \hat{p}_i| \xrightarrow{P} 0, \quad \text{as } T_i, B \rightarrow \infty, \quad (5)$$

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.

## 2.1 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

$$\begin{aligned} y_{i,t} &= \left( \eta_i + \sum_{j=1}^{q_1} \phi_{i,j} y_{i,t-j} + \sum_{j=0}^{q_2-1} \theta'_{i,j+1} \mathbf{x}_{i,t-j} \right) (1 - D_{i,t}) + f(\mathcal{F}_{i,t}, \alpha_i) D_{i,t} + \varepsilon_{i,t}, \\ \mathcal{M}: \quad f(\mathcal{F}_{i,t}, \alpha_i) &= \alpha_i + \sum_{j=1}^{q_1} \tilde{\phi}_{i,j} y_{i,t-j} + \sum_{j=0}^{q_2-1} \tilde{\theta}'_{i,j+1} \mathbf{x}_{i,t-j}, \\ \alpha_i &= \mu_\alpha + \delta'_i \mathbf{x}_{i,T_i^*+1} + \varepsilon_{\alpha,i}, \end{aligned} \quad (6)$$

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  $\phi_i = (\phi_{i,1}, \dots, \phi_{i,q_1})'$ ,  $\theta_i = (\theta_{i,1}, \dots, \theta_{i,q_2})'$ ,  $\tilde{\phi}_i = (\tilde{\phi}_{i,1}, \dots, \tilde{\phi}_{i,q_1})'$ ,  $\tilde{\theta}_i = (\tilde{\theta}_{i,1}, \dots, \tilde{\theta}_{i,q_2})'$ ,  $\delta_i = (\delta_{i,1}, \dots, \delta_{i,p})'$ , and suppose that the regression coefficients in (6) have the following hierarchical random effects structure:

$$\begin{aligned}
\eta_i &\stackrel{iid}{\sim} \mathcal{F}_\eta \text{ with } E_{\mathcal{F}_\eta}(\eta_i) = \mu_\eta, \text{Var}_{\mathcal{F}_\eta}(\eta_i) = \sigma_\eta^2, \\
\phi_{i,j} &\stackrel{iid}{\sim} \mathcal{F}_{\phi_j} \text{ where } |\phi_{i,j}| < 1, \\
\theta_{i,j} &\stackrel{iid}{\sim} \mathcal{F}_{\theta_j} \text{ with } E_{\mathcal{F}_{\theta_j}}(\theta_{i,j}) = \mu_{\theta_j}, \text{Var}_{\mathcal{F}_{\theta_j}}(\theta_{i,j}) = \Sigma_{\theta_j}^2, \\
\delta_i &\stackrel{iid}{\sim} \mathcal{F}_\delta \text{ with } E_{\mathcal{F}_\delta}(\delta_i) = \mu_\delta, \text{Var}_{\mathcal{F}_\delta}(\delta_i) = \Sigma_\delta, \\
\varepsilon_{i,t} &\stackrel{iid}{\sim} \mathcal{F}_{\varepsilon_i} \text{ with } E_{\mathcal{F}_{\varepsilon_i}}(\varepsilon_{i,t}) = 0, \text{Var}_{\mathcal{F}_{\varepsilon_i}}(\varepsilon_{i,t}) = \sigma_i^2, \\
\varepsilon_{\alpha,i} &\stackrel{iid}{\sim} \mathcal{F}_{\varepsilon_\alpha} \text{ with } E_{\mathcal{F}_{\varepsilon_\alpha}}(\varepsilon_{\alpha,i}) = 0, \text{Var}_{\mathcal{F}_{\varepsilon_\alpha}}(\varepsilon_{\alpha,i}) = \sigma_\alpha^2, \\
\tilde{\phi}_{i,j} &\stackrel{ind}{\sim} \mathcal{F}_{\tilde{\phi}_j}(\mathbf{x}_{i,T_i^*+1}) \text{ where } |\tilde{\phi}_{i,j}| < 1, \\
\tilde{\theta}_{i,j} &\stackrel{ind}{\sim} \mathcal{F}_{\tilde{\theta}_j}(\mathbf{x}_{i,T_i^*+1}), \\
\eta_i &\perp\!\!\!\perp \phi_{i,j} \perp\!\!\!\perp \theta_{i,j} \perp\!\!\!\perp \tilde{\phi}_{i,j} \perp\!\!\!\perp \lambda_{i,j} \perp\!\!\!\perp \delta_i \perp\!\!\!\perp \varepsilon_{i,t} \perp\!\!\!\perp \tilde{\varepsilon}_i,
\end{aligned} \tag{7}$$

where the distributions  $\mathcal{F}_{\tilde{\phi}_j}(\mathbf{x}_{i,T_i^*+1})$  and  $\mathcal{F}_{\tilde{\theta}_j}(\mathbf{x}_{i,T_i^*+1})$  satisfy  $\mathcal{F}_{\tilde{\phi}_j}(\mathbf{x}_{i,T_i^*+1}) \stackrel{d}{=} \mathcal{F}_{\tilde{\phi}_j}(\mathbf{x}_{i',T_i^*+1})$  and  $\mathcal{F}_{\tilde{\theta}_j}(\mathbf{x}_{i,T_i^*+1}) \stackrel{d}{=} \mathcal{F}_{\tilde{\theta}_j}(\mathbf{x}_{i',T_i^*+1})$  when  $\mathbf{x}_{i,T_i^*+1} = \mathbf{x}_{i',T_i^*+1}$ . Note that for model (6) 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.

We see that model (6) with its accompanying random effects structure (7) is flexible enough to capture changing structural dynamics as well as a mean-shift, and is a generalization of the model 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.

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  $\|\mathbf{x}_{i,T_i^*+1} - \mathbf{x}_{j,T_j^*+1}\|_2$  are expected to experience similar structural changes. This makes distance based weighting an attractive avenue.

## 2.2 Forecasting and testing for shock persistence

In this section, we motivate the relation between shock persistence testing and forecasting comparison. In time series literature, under a squared error loss  $L(\cdot)$ , the forecast using the conditional mean is the best point forecast that minimizes the expected loss if the model is specified correctly. That being said, even if the time series does experience a shock but the shock is not persistent, the expected loss of the conditional mean forecast should converge to the minimum expected loss that any forecast can achieve as sample size increases. On the other hand, if the shock is persistent, the conditional mean forecast that incorporates the shock information and makes good use of it should enjoy a smaller expected loss than any conditional mean forecast without such information does. In other words, the problem of testing for shock persistence boils down into comparing the conditional mean forecast that is adjusted by shock information and the one that is unadjusted. To deal with this problem, we applied the forecast comparison procedure of [Quaedvlieg \[2021\]](#) to conduct shock persistence testing.

In this paper, we extend the framework of post-shock forecast [[Lin and Eck, 2021](#)], which aggregates outside information by constructing a donor pool, to post-shock persistence testing. The donor pool consists of time series called “donors” that are similar to the time series of interest. As we do not observe the future data, information from donors can aid our decision making and forecasts. Note that what similarity exactly means here can be vague but the donors and the time series of interest should come

from a common population in principle; see more details of how to select donors in [Lin and Eck \[2021\]](#). In this paper, we also proposed several theoretical results with some detailed conditions on the donor pool, which may alleviate the difficulty of dealing with similarity when choosing donors.

In this paper, we propose a new way of aggregating information from donor pool to conduct post-shock persistence testing, which is quite different from [Lin and Eck \[2021\]](#). The general structure of this aggregation technique is illustrated as below. Suppose we are given a nonempty donor pool with size  $n$  that consists of donors  $\mathcal{T}_2, \dots, \mathcal{T}_{n+1}$ , whose data, including shock information, are observed. With shock information, we are able to compute unadjusted forecast that does not incorporate shock information and adjusted forecast that does. The forecast comparison using methodology of [Quaedvlieg \[2021\]](#) for each donor yields a  $p$ -value,  $p_{i+1}$  for  $i$ th donor in the donor pool. Given  $\{p_i: i = 2, \dots, n+1\}$ , we propose a methodology for judging whether the shock of the time series of interest is persistent or not. Next, we illustrate the details of our methodology as below.

For each donor  $\mathcal{T}_i$ ,  $i = 2, \dots, n+1$ , to compare adjusted and unadjusted forecasts, a model needs to be specified. The model should describe how the shock impacts the time series at and after a certain time point  $t = T_i^* + 1$ . For clarity, we refer the unadjusted forecast to the conditional mean forecast based on the part of the model that ignores shock effect; and the adjusted forecast refers to the conditional mean forecast based on the full model. Note that the model needs not to be the same for every donor in the donor pool. However, certain conditions need to be imposed as discussed previously.

Given a model for each donor pool, we compute the unadjusted conditional mean forecast and adjusted conditional mean forecast in the framework of [Quaedvlieg \[2021\]](#). As introduced in the beginning of Section 2, for a given horizon  $h$ , to forecast  $y_{i,t}$ , we compute the two forecasts based on the information set  $\mathcal{F}_{i,t-h}$  of each donor pool (i.e., the data before  $t-h$ ) for  $i = 2, \dots, n+1$ . Note that in practice, a training sample size  $K_i$  should be specified for  $i = 2, \dots, n+1$  such that we use the information set  $\mathcal{F}_{i,t-h-K_i+1:t-h}$ .

Even if the data contain shock information for each donor, it is impossible to compute the adjusted conditional mean forecast based on  $\mathcal{F}_{i,t-h}$  when  $t-h \leq T_i^*$ , i.e., we only have training data that do not have any shock information. Note that this is not a big problem when  $t-h \geq T_i^* + 1$  because  $\mathcal{F}_{i,t-h}$  starts to incorporate shock information. In the literature, [Lin and Eck \[2021\]](#) utilizes donor pool information to provide a consistent estimator of the shock effect such that the adjusted conditional mean forecast is the unadjusted one plus the shock effect estimate. However, the consistent estimation comes with the cost of assuming certain models (in terms of first and second moments) for the shock effect across the donor pool and the method only works for additive shock effect model. To our best knowledge, there is no other available methods that can accommodate the case when the shock model is nonlinear. However, even if the model is misspecified, using additive shock effect model to approximate the scalar shock effect that occurs at  $t = T_i^* + 1$  may still help as opposed to doing nothing.

With the presence of autoregressive parameters in some models, unbiased estimation with OLS is not feasible but the bias is  $O(T^{-1})$ , where  $T$  is the time series length [[Shaman and Stine, 1988](#)]. In this paper, we use the OLS estimates to approximate the adjusted conditional mean forecast and unadjusted one. The bias decays away as  $T \rightarrow \infty$ .

**Note:** Our forecast needs to be written with respect to our general model. Specifics can be given when we conduct our numerical examples.

In our post-shock setting we consider the following candidate forecasts:

Forecast 1 :

Forecast 2 :

We want to determine which forecast is appropriate over a horizon while the methods in [Lin and Eck \[2021\]](#) were only appropriate in the nowcasting setting in which prediction was only focused on the response immediately following the shock.

**Note:** We need to explain what the voting method is, possibly in algorithmic format. Also note that the p-values that we obtain are computed using a bootstrap procedure. Perhaps an additional proposition that states the performance of these bootstrap p-values would further guarantee reliability.

**Proposition 1.** Suppose  $p_2, \dots, p_{n+1}$  is a sequence of pairwise independent p-values with  $\mathbb{P}(p_i \leq \alpha) = \kappa$  for  $i = 1, \dots, n+1$ , where  $\alpha$  is the significance level,  $\kappa$  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  $n \rightarrow \infty$ ,

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

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

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

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

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

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

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

That is, the expected misclassification rate of voting converges to

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

□

**Lemma 1.** For each  $n \in \mathbb{N}$ , let  $c_{n1}, \dots, c_{nn}$  be real numbers bounded by some  $K > 0$ , and let  $X_{n1}, \dots, X_{nn}$  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

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

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

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

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

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

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.,  $\mathbf{Var}[T_n]/(b_n)^2 \rightarrow 0$ . Since  $X_{n1}, \dots, X_{nn}$  are pairwise independent, so are  $Y_{n1}, \dots, Y_{nn}$ , and consequently

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

As a result,

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

which proves the result by sandwich theorem.  $\square$

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

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

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

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

*where  $\kappa_i = \mathbb{P}(p_i \leq \alpha) \in [0, 1]$  for  $i \in \mathcal{I}_n$ ,  $\alpha$  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  $n \rightarrow \infty$ ,*

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

*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,  $|\mathcal{I}_n| \rightarrow \infty$ ,  $b_n > 1$ , and  $b_n \rightarrow \infty$  as  $n \rightarrow \infty$ ,

$$\begin{aligned} \sum_{i \in \mathcal{I}_n} \mathbb{P}(I(p_i \leq \alpha) > b_n) &\rightarrow 0 \\ \frac{1}{b_n^2} \sum_{i \in \mathcal{I}_n} \mathbb{E}[I(p_i \leq \alpha) | I(p_i \leq \alpha) \leq b_n] &= \frac{1}{b_n^2} \sum_{i \in \mathcal{I}_n} \mathbb{P}(p_i \leq \alpha) = \frac{1}{b_n^2} \sum_{i \in \mathcal{I}_n} w_i \kappa_i \rightarrow 0 \end{aligned}$$

because  $\sum_{i \in \mathcal{I}_n} w_i \kappa_i = O(1)$  by the assumption  $\sum_{i \in \mathcal{I}_n} w_i \kappa_i \rightarrow \kappa_1$ . 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

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



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  $\sum_{i \in \mathcal{I}_n} w_i \kappa_i \rightarrow \kappa_1$  as  $n \rightarrow \infty$ , 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.  $\square$

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  $n \rightarrow \infty$ , where  $\mathcal{I}_n = \{i = 2, \dots, n+1 : 0 < w_i < 1\}$ . 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 Corollary 1 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

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

where the last second step assumes that  $\{I(p_i \leq \alpha) : i \in \mathcal{I}_n\}$  are independent and the last step uses the condition  $\sum_{i \in \mathcal{I}_n} w_i \kappa_i \rightarrow \kappa_1$  as  $n \rightarrow \infty$ , which is used in Corollary 1. Notice that if  $\kappa_i$  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 Corollary 1 should be acceptable.

Suppose we are interested in comparing two models  $\mathcal{M}_1$  and  $\mathcal{M}_2$  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  $i$ th time series be  $a_i^j$  for  $i = 1, \dots, n+1$  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  $\mathcal{M}_1$  and is true for all the donors and time series of interest,  $\xi_i$  is expected to be high. Additionally, it is reasonable to assume  $\xi_i = \xi$  for  $\xi \in [0, 1]$  and  $i = 1, \dots, n+1$ . 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 2** (Synthetic pairwise model selection). *Let  $(w_2, \dots, w_{n+1})$  be weights such that  $w_i \in [0, 1]$  and  $\sum_{i=2}^{n+1} w_i = 1$ . Define*

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

*Suppose that  $\mathcal{I}_n$  is non-empty,  $|\mathcal{I}_n| \rightarrow \infty$  as  $n \rightarrow \infty$ , and  $w_i b_n \leq K$  for  $i \in \mathcal{I}_n$  and some  $K > 0$ , where  $b_n \geq 1$  and  $b_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Assume for  $i \in \mathcal{I}_n$ ,  $\{a_i^1, a_i^2\}$  are AICs of  $\mathcal{M}_1$  and  $\mathcal{M}_2$  for  $i$ th time series, which are pairwise independent of  $\{a_j^1, a_j^2\}$  for  $j \neq i$  and that*

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

*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  $n \rightarrow \infty$ ,*

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

Note that Corollary 2 gives an asymptotic upper bound for the misclassification rate of synthetic pairwise model selection, which is small if  $\xi_1$  is large.

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 (6) and (7) 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  $\Sigma_{\theta_k}$ ,  $\Sigma_\delta$ , and  $\Sigma_\gamma$  for  $k = 1, \dots, q_2$ . Define

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

**Proposition 2.** *Let  $\mathbf{y}_{i,t}$ ,  $i = 1, \dots, n_i$ ,  $t = 1, \dots, T$  be a collection of independent time series. Let each time series  $\mathbf{y}_{i,t}$  be generated by model (6) with accompanying random effects structure (7), where  $y_{i,t}$  is a scalar response  $\mathbf{x}_{i,t} \in \mathbb{R}^p$  is a fixed vector of covariates. Suppose that there is a common shock time  $p+2 < T^* < T - (p+2)$  where  $q_1 = q_2 = 1$ ,  $\|\mathbf{x}_{i,t}\| = 1$ , each entry of  $\mathbf{x}_{i,t}$  is non-negative,  $\mu_{\min} > 0$ , and  $y_{i,0} = 0$ . Then for every  $1 \leq t \leq T$  we have that*

$$\mathbb{E}(y_{i,t}) \geq C \text{Var}(y_{i,t}), \quad (8)$$

provided that

$$\begin{aligned}\mu_\alpha + 3\mu_{\min} &\geq C(3\sigma_{\max}^2 + 2\lambda_{\max}), \\ 2\mu_{\min} &\geq C(2\sigma_{\max}^2 + \lambda_{\max}).\end{aligned} \quad (9)$$

**Remarks:**

1. The conditions (9) govern the signal strength necessary in the expected behavior of the underlying random effects structure relative to variability necessary for the overall signal to noise ratio for the overall time series to be lower bounded by some positive value  $C$ .
2. The conditions of Proposition 2 state that there is an increasing trend in  $y_{i,t}$  for all  $i$ . This can be relaxed, but a fully general argument is cumbersome.

*Proof.* For  $t > T^* + 1$  we have

$$\begin{aligned}\mathbb{E}(y_{i,t}) &= \mathbb{E}[\mathbb{E}(y_{i,t} | y_{i,t-1})] \\ &= \mathbb{E}[\mathbb{E}(\eta_i + \alpha_i + \tilde{\phi}_i y_{i,t-1} + \tilde{\theta}_i' \mathbf{x}_{i,t} | y_{i,t-1})] \\ &= \mathbb{E}(\eta_i + \mu_\alpha + \delta_i' \mathbf{x}_{i,T^*+1} + \tilde{\phi}_i y_{i,t-1} + \tilde{\theta}_i' \mathbf{x}_{i,t}) \\ &= \mu_\eta + \mu_\alpha + \mu_\delta' \mathbf{x}_{i,T^*+1} + \mu_{\tilde{\theta}}' \mathbf{x}_{i,t} + \mu_{\tilde{\phi}} \mathbb{E}(y_{i,t-1}) \\ &\geq (\mu_\alpha + 3\mu_{\min}) + \mu_{\min} \mathbb{E}(y_{i,t-1}) \\ &\geq \sum_{j=1}^{t-(T^*+1)} \mu_{\min}^j (\mu_\alpha + 3\mu_{\min}) + \mu_{\min}^{t-T^*} \mathbb{E}(y_{i,T^*}) \\ &= \sum_{j=0}^{t-(T^*+1)} \mu_{\min}^j (\mu_\alpha + 3\mu_{\min}) + \mu_{\min}^{t-T^*} (\mathbb{E}[\mathbb{E}(y_{i,T^*} | y_{i,T^*-1})])\end{aligned}$$

$$\begin{aligned}
&= \sum_{j=0}^{t-(T^*+1)} \mu_{\min}^j (\mu_\alpha + 3\mu_{\min}) + \mu_{\min}^{t-T^*} (\mathbb{E}(\eta_i + \theta'_i \mathbf{x}_{i,T^*} + \phi_i y_{i,T^*-1})) \\
&\geq \sum_{j=0}^{t-(T^*+1)} \mu_{\min}^j (\mu_\alpha + 3\mu_{\min}) + \mu_{\min}^{t-T^*} (2\mu_{\min} + \mu_{\min} \mathbb{E}(y_{i,T^*-1})) \\
&\geq \sum_{j=0}^{t-(T^*+1)} \mu_{\min}^j (\mu_\alpha + 3\mu_{\min}) + \mu_{\min}^{t-T^*} \left( \sum_{j=0}^{T^*-1} \mu_{\min}^j (2\mu_{\min}) \right),
\end{aligned}$$

and

$$\begin{aligned}
\text{Var}(y_{i,t}) &= \mathbb{E} [\text{Var}(y_{i,t} | y_{i,t-1})] + \text{Var} [\mathbb{E}(y_{i,t} | y_{i,t-1})] \\
&= \mathbb{E} \left[ \text{Var}(\eta_i + \mu_\alpha + \delta'_i \mathbf{x}_{i,T^*+1} + \tilde{\theta}'_i \mathbf{x}_{i,t} + \tilde{\phi}_i y_{i,t-1} + \varepsilon_{i,t} + \varepsilon_{\alpha,i} | y_{i,t-1}) \right] \\
&\quad + \text{Var} \left[ \mathbb{E}(\eta_i + \mu_\alpha + \delta'_i \mathbf{x}_{i,T^*+1} + \tilde{\theta}'_i \mathbf{x}_{i,t} + \tilde{\phi}_i y_{i,t-1} + \varepsilon_{i,t} + \varepsilon_{\alpha,i} | y_{i,t-1}) \right] \\
&\leq 3\sigma_{\max}^2 + 2\lambda_{\max} + \text{Var} \left[ \mathbb{E}(\tilde{\phi}_i y_{i,t-1} | y_{i,t-1}) \right] \\
&= 3\sigma_{\max}^2 + 2\lambda_{\max} + \mu_{\tilde{\phi}}^2 \text{Var}(y_{i,t-1}) \\
&\leq \sum_{j=0}^{t-(T^*+1)} (\mu_{\tilde{\phi}}^2)^j (3\sigma_{\max}^2 + 2\lambda_{\max}) + (\mu_{\tilde{\phi}}^2)^{t-T^*} \text{Var}(y_{i,T^*}) \\
&= \sum_{j=0}^{t-(T^*+1)} (\mu_{\tilde{\phi}}^2)^j (3\sigma_{\max}^2 + 2\lambda_{\max}) + (\mu_{\tilde{\phi}}^2)^{t-T^*} (\mathbb{E} [\text{Var}(y_{i,T^*} | y_{i,T^*-1})] + \text{Var} [\mathbb{E}(y_{i,T^*} | y_{i,T^*-1})]) \\
&= \sum_{j=0}^{t-(T^*+1)} (\mu_{\tilde{\phi}}^2)^j (3\sigma_{\max}^2 + 2\lambda_{\max}) + (\mu_{\tilde{\phi}}^2)^{t-T^*} (\mathbb{E} [\text{Var}(\eta_i + \theta'_i \mathbf{x}_{i,T^*} + \phi_i y_{i,T^*-1} + \varepsilon_{i,T^*} | y_{i,T^*-1})] \\
&\quad + \text{Var} [\mathbb{E}(\eta_i + \theta'_i \mathbf{x}_{i,T^*} + \phi_i y_{i,T^*-1} + \varepsilon_{i,T^*} | y_{i,T^*-1})]) \\
&\leq \sum_{j=0}^{t-(T^*+1)} (\mu_{\tilde{\phi}}^2)^j (3\sigma_{\max}^2 + 2\lambda_{\max}) + (\mu_{\tilde{\phi}}^2)^{t-T^*} (2\sigma_{\max}^2 + \lambda_{\max} + \mu_{\tilde{\phi}}^2 \text{Var}(y_{i,T^*-1})) \\
&\leq \sum_{j=0}^{t-(T^*+1)} (\mu_{\tilde{\phi}}^2)^j (3\sigma_{\max}^2 + 2\lambda_{\max}) + (\mu_{\tilde{\phi}}^2)^{t-T^*} \left( \sum_{j=0}^{T^*-1} (\mu_{\tilde{\phi}}^2)^j (2\sigma_{\max}^2 + \lambda_{\max}) \right).
\end{aligned}$$

Combining the above with the condition (9) yields

$$\begin{aligned}
\frac{\mathbb{E}(y_{i,t})}{\text{Var}(y_{i,t})} &\geq \frac{\sum_{j=0}^{t-(T^*+1)} \mu_{\min}^j (\mu_\alpha + 3\mu_{\min}) + \mu_{\min}^{t-T^*} \left( \sum_{j=0}^{T^*-1} \mu_{\min}^j (2\mu_{\min}) \right)}{\sum_{j=0}^{t-(T^*+1)} (\mu_{\tilde{\phi}}^2)^j (3\sigma_{\max}^2 + 2\lambda_{\max}) + (\mu_{\tilde{\phi}}^2)^{t-T^*} \left( \sum_{j=0}^{T^*-1} (\mu_{\tilde{\phi}}^2)^j (2\sigma_{\max}^2 + \lambda_{\max}) \right)} \\
&\geq \frac{C \sum_{j=0}^{t-(T^*+1)} (\mu_{\tilde{\phi}}^2)^j (3\sigma_{\max}^2 + 2\lambda_{\max}) + C(\mu_{\tilde{\phi}}^2)^{t-T^*} \left( \sum_{j=0}^{T^*-1} (\mu_{\tilde{\phi}}^2)^j (2\sigma_{\max}^2 + \lambda_{\max}) \right)}{\sum_{j=0}^{t-(T^*+1)} (\mu_{\tilde{\phi}}^2)^j (3\sigma_{\max}^2 + 2\lambda_{\max}) + (\mu_{\tilde{\phi}}^2)^{t-T^*} \left( \sum_{j=0}^{T^*-1} (\mu_{\tilde{\phi}}^2)^j (2\sigma_{\max}^2 + \lambda_{\max}) \right)} \\
&= C,
\end{aligned}$$

for all  $i$  and  $t$ . A similar argument holds when  $t \leq T^* + 1$ .  $\square$

**Lemma 2** (Technical Lemma 1). *Let  $a_j, b_j \in \mathbb{R}$ ,  $j = 1, 2$  be such that  $0 < a_1, a_2 < 1$ ,  $b_1, b_2 > 0$ . Let  $c_1 = \max(a_1, a_2)$ ,  $c_2 = \min(a_1, a_2)$ ,  $d_1 = \max(b_1, b_2)$ ,  $d_2 = \min(b_1, b_2)$ . Then,*

$$|a_1 b_1 - a_2 b_2| \leq |a_1 - a_2| d_1 + |b_1 - b_2| c_2.$$

*Proof.* Observe that

$$|a_1 b_1 - a_2 b_2| \leq c_1 d_1 - c_2 d_2 = (c_1 - c_2) d_1 + (d_1 - d_2) c_2.$$

The conclusion follows by noting that  $c_1 - c_2 = |a_1 - a_2|$  and  $d_1 - d_2 = |b_1 - b_2|$ .  $\square$

**Proposition 3.** *Let  $\mathbf{y}_{i,t}$ ,  $i = 1, \dots, n_i$ ,  $t = 1, \dots, T$  be a collection of independent time series. Let each time series  $\mathbf{y}_{i,t}$  be generated by model (6) with accompanying random effects structure (7), where  $y_{i,t}$  is a scalar response  $\mathbf{x}_{i,t} \in \mathbb{R}^p$  is a fixed vector of covariates. Suppose that there is a common shock time  $p + 2 < T^* < T - (p + 2)$  where  $q_1 = q_2 = 1$ ,  $\|\mathbf{x}_{i,t}\| = 1$ , each entry of  $\mathbf{x}_{i,t}$  is non-negative,  $\mu_{\min} > 0$ , and  $y_{i,0} = 0$ . Let  $i' \geq 2$  be an index. Then, for any  $\epsilon > 0$  we can choose an  $\gamma > 0$  such that*

$$|\mathbb{E}(y_{i,t} - y_{i',t})| \leq \epsilon, \quad |\text{Var}(y_{1,t}) - \text{Var}(y_{i',t})| < \epsilon,$$

*provided that*

$$\begin{aligned} |\mu'_{\delta} \mathbf{x}_{1,T^*+1} - \mu'_{\delta} \mathbf{x}_{i',T^*+1}| &< \gamma, \\ \left| \mu'_{\tilde{\theta}_1} \mathbf{x}_{1,t} - \mu'_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t} \right| &< \gamma, \quad \text{for all } t, \\ |\mu'_{\tilde{\phi}_1} - \mu'_{\tilde{\phi}_{i'}}| &< \gamma, \quad \text{for all } j, \\ |\mathbf{x}'_{1,T^*+1} \Sigma_{\delta} \mathbf{x}_{1,T^*+1} - \mathbf{x}'_{i',T^*+1} \Sigma_{\delta} \mathbf{x}_{i',T^*+1}| &< \gamma, \\ |\mathbf{x}'_{1,t} \Sigma_{\tilde{\theta}_1} \mathbf{x}_{1,t} - \mathbf{x}'_{i',t} \Sigma_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t}| &< \gamma, \quad \text{for all } t. \end{aligned}$$

*Proof.* A similar recursion to that seen in the proof of Proposition 2 yields

$$\begin{aligned} |\mathbb{E}(y_{1,t} - y_{i',t})| &= \left| \sum_{j=0}^{t-(T^*+1)} \left[ \left( \mu_{\tilde{\phi}_1}^j (\mu'_{\delta} \mathbf{x}_{1,T^*+1}) - \mu_{\tilde{\phi}_{i'}}^j (\mu'_{\delta} \mathbf{x}_{i',T^*+1}) \right) + \mu_{\tilde{\phi}_1}^j (\mu'_{\tilde{\theta}_1} \mathbf{x}_{1,t-j}) - \mu_{\tilde{\phi}_{i'}}^j (\mu'_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t-j}) \right] \right. \\ &\quad \left. + \sum_{j=0}^{T^*-1} \left[ \mu_{\eta} (\mu_{\tilde{\phi}_1}^{t-T^*} - \mu_{\tilde{\phi}_{i'}}^{t-T^*}) \mu_{\phi}^j + \left( \mu_{\tilde{\phi}_1}^{t-T^*} (\mu'_{\theta} \mathbf{x}_{1,t-j}) - \mu_{\tilde{\phi}_{i'}}^{t-T^*} (\mu'_{\theta} \mathbf{x}_{i',t-j}) \right) \mu_{\phi}^j \right] \right| \\ &\leq |\mu'_{\delta} \mathbf{x}_{1,T^*+1} - \mu'_{\delta} \mathbf{x}_{i',T^*+1}| + \sum_{j=1}^{t-(T^*+1)} \left| \mu_{\tilde{\phi}_1}^j (\mu'_{\delta} \mathbf{x}_{1,T^*+1}) - \mu_{\tilde{\phi}_{i'}}^j (\mu'_{\delta} \mathbf{x}_{i',T^*+1}) \right| \\ &\quad + |\mu'_{\tilde{\theta}_1} \mathbf{x}_{1,t} - \mu'_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t}| + \sum_{j=1}^{t-(T^*+1)} \left| \mu_{\tilde{\phi}_1}^j (\mu'_{\tilde{\theta}_1} \mathbf{x}_{1,t-j}) - \mu_{\tilde{\phi}_{i'}}^j (\mu'_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t-j}) \right| \\ &\quad + \mu_{\eta} \sum_{j=0}^{T^*-1} \left| \mu_{\tilde{\phi}_1}^{t-T^*} - \mu_{\tilde{\phi}_{i'}}^{t-T^*} \right| \mu_{\phi}^j + \sum_{j=0}^{T^*-1} \left| \mu_{\tilde{\phi}_1}^{t-T^*} (\mu'_{\theta} \mathbf{x}_{1,T^*-j}) - \mu_{\tilde{\phi}_{i'}}^{t-T^*} (\mu'_{\theta} \mathbf{x}_{i',T^*-j}) \right| \mu_{\phi}^j \\ &\leq |\mu'_{\delta} \mathbf{x}_{1,T^*+1} - \mu'_{\delta} \mathbf{x}_{i',T^*+1}| + |\mu'_{\tilde{\theta}_1} \mathbf{x}_{1,t} - \mu'_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t}| \\ &\quad + \sum_{j=1}^{t-(T^*+1)} \left[ |\mu_{\tilde{\phi}_1}^j - \mu_{\tilde{\phi}_{i'}}^j| \max(\mu'_{\delta} \mathbf{x}_{1,T^*+1}, \mu'_{\delta} \mathbf{x}_{i',T^*+1}) + |\mu'_{\delta} \mathbf{x}_{1,T^*+1} - \mu'_{\delta} \mathbf{x}_{i',T^*+1}| \mu_{\tilde{\phi}_1}^j \right] \\ &\quad + \sum_{j=1}^{t-(T^*+1)} \left[ |\mu_{\tilde{\phi}_1}^j - \mu_{\tilde{\phi}_{i'}}^j| \max(\mu'_{\tilde{\theta}_1} \mathbf{x}_{1,t-j}, \mu'_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t-j}) + |\mu'_{\tilde{\theta}_1} \mathbf{x}_{1,t-j} - \mu'_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t-j}| \mu_{\tilde{\phi}_1}^j \right] \end{aligned}$$

$$+ \mu_\eta \sum_{j=0}^{T^*-1} \left| \mu_{\tilde{\phi}_1}^{t-T^*} - \mu_{\tilde{\phi}_{i'}}^{t-T^*} \right| \mu_\phi^j + \sum_{j=0}^{T^*-1} \left| \mu_{\tilde{\phi}_1}^{t-T^*} (\mu'_\theta \mathbf{x}_{1,T^*-j}) - \mu_{\tilde{\phi}_{i'}}^{t-T^*} (\mu'_\theta \mathbf{x}_{i',T^*-j}) \right| \mu_\phi^j,$$

where the second inequality follows from applications of Lemma 2. Continuing from the second inequality, the conditions of this proposition imply that

$$\begin{aligned} & |\mathbb{E}(y_{1,t} - y_{i',t})| \\ & \leq 2\gamma + \sum_{j=1}^{\infty} \left[ \left| \mu_{\tilde{\phi}_1}^j - \mu_{\tilde{\phi}_{i'}}^j \right| \max(\mu'_\delta \mathbf{x}_{1,T^*+1}, \mu'_\delta \mathbf{x}_{i',T^*+1}) + \gamma \mu_{\tilde{\phi}_1}^j \right] \\ & \quad + \sum_{j=1}^{\infty} \left[ \left| \mu_{\tilde{\phi}_1}^j - \mu_{\tilde{\phi}_{i'}}^j \right| \max(\mu'_\delta \mathbf{x}_{1,t-j}, \mu'_\delta \mathbf{x}_{i',t-j}) + \gamma \mu_{\tilde{\phi}_1}^j \right] \\ & \quad + \mu_\eta \sum_{j=0}^{\infty} \left| \mu_{\tilde{\phi}_1}^{t-T^*} - \mu_{\tilde{\phi}_{i'}}^{t-T^*} \right| \mu_\phi^j + \sum_{j=0}^{\infty} \left| \mu_{\tilde{\phi}_1}^{t-T^*} (\mu'_\theta \mathbf{x}_{1,T^*-j}) - \mu_{\tilde{\phi}_{i'}}^{t-T^*} (\mu'_\theta \mathbf{x}_{i',T^*-j}) \right| \mu_\phi^j \\ & \leq 2\gamma + 2\gamma \frac{\max(\mu_\delta)}{(1 - \mu_{\tilde{\phi}_1})(1 - \mu_{\tilde{\phi}_{i'}})} \\ & \quad + \mu_\eta \sum_{j=0}^{\infty} \left| \mu_{\tilde{\phi}_1}^{t-T^*} - \mu_{\tilde{\phi}_{i'}}^{t-T^*} \right| \mu_\phi^j + \sum_{j=0}^{\infty} \left| \mu_{\tilde{\phi}_1}^{t-T^*} (\mu'_\theta \mathbf{x}_{1,T^*-j}) - \mu_{\tilde{\phi}_{i'}}^{t-T^*} (\mu'_\theta \mathbf{x}_{i',T^*-j}) \right| \mu_\phi^j \\ & \leq 2\gamma + 2\gamma \frac{\max(\mu_\delta)}{(1 - \mu_{\tilde{\phi}_1})(1 - \mu_{\tilde{\phi}_{i'}})} + \gamma \frac{\mu_\eta}{1 - \mu_\phi} + \sum_{j=0}^{\infty} \left| \mu_{\tilde{\phi}_1}^{t-T^*} (\mu'_\theta \mathbf{x}_{1,T^*-j}) - \mu_{\tilde{\phi}_{i'}}^{t-T^*} (\mu'_\theta \mathbf{x}_{i',T^*-j}) \right| \mu_\phi^j \\ & \leq 2\gamma + 2\gamma \frac{\max(\mu_\delta)}{(1 - \mu_{\tilde{\phi}_1})(1 - \mu_{\tilde{\phi}_{i'}})} + \gamma \frac{\mu_\eta}{1 - \mu_\phi} + \gamma \frac{\max(\mu_\theta) + 1}{1 - \mu_\phi}, \end{aligned}$$

where the last inequality follows from Lemma 2. We can choose  $\gamma$  small enough so that  $|\mathbb{E}(y_{1,t} - y_{i',t})| < \varepsilon$ . A similar recursion for the variance yields

$$\begin{aligned} & |\text{Var}(y_{1,t}) - \text{Var}(y_{i',t})| = \left| \mathbf{x}'_{1,T^*+1} \Sigma_\delta \mathbf{x}_{1,T^*+1} - \mathbf{x}'_{i',T^*+1} \Sigma_\delta \mathbf{x}_{i',T^*+1} + \mathbf{x}'_{1,t} \Sigma_{\tilde{\theta}_1} \mathbf{x}_{1,t} - \mathbf{x}'_{i',t} \Sigma_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t} \right. \\ & \quad \left. + \mu_{\tilde{\phi}_1}^2 \text{Var}(y_{1,t-1}) - \mu_{\tilde{\phi}_{i'}}^2 \text{Var}(y_{i',t-1}) \right| \\ & \leq \left| \mathbf{x}'_{1,T^*+1} \Sigma_\delta \mathbf{x}_{1,T^*+1} - \mathbf{x}'_{i',T^*+1} \Sigma_\delta \mathbf{x}_{i',T^*+1} + \mathbf{x}'_{1,t} \Sigma_{\tilde{\theta}_1} \mathbf{x}_{1,t} - \mathbf{x}'_{i',t} \Sigma_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t} \right| \\ & \quad + \left| \mu_{\tilde{\phi}_1}^2 \text{Var}(y_{1,t-1}) - \mu_{\tilde{\phi}_{i'}}^2 \text{Var}(y_{i',t-1}) \right| \\ & = \left| \mathbf{x}'_{1,T^*+1} \Sigma_\delta \mathbf{x}_{1,T^*+1} - \mathbf{x}'_{i',T^*+1} \Sigma_\delta \mathbf{x}_{i',T^*+1} + \mathbf{x}'_{1,t} \Sigma_{\tilde{\theta}_1} \mathbf{x}_{1,t} - \mathbf{x}'_{i',t} \Sigma_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t} \right| \\ & \quad + \left| \mu_{\tilde{\phi}_1}^2 \left( \sigma_\eta^2 + \sigma_\alpha^2 + \sigma^2 + \mathbf{x}'_{1,T^*+1} \Sigma_\delta \mathbf{x}_{1,T^*+1} + \mathbf{x}'_{1,t-1} \Sigma_{\tilde{\theta}_1} \mathbf{x}_{1,t-1} + \mu_{\tilde{\phi}_1}^2 \text{Var}(y_{1,t-2}) \right) \right. \\ & \quad \left. - \mu_{\tilde{\phi}_{i'}}^2 \left( \sigma_\eta^2 + \sigma_\alpha^2 + \sigma^2 + \mathbf{x}'_{i',T^*+1} \Sigma_\delta \mathbf{x}_{i',T^*+1} + \mathbf{x}'_{i',t-1} \Sigma_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t-1} + \mu_{\tilde{\phi}_{i'}}^2 \text{Var}(y_{i',t-2}) \right) \right| \\ & \leq \left| \mathbf{x}'_{1,T^*+1} \Sigma_\delta \mathbf{x}_{1,T^*+1} - \mathbf{x}'_{i',T^*+1} \Sigma_\delta \mathbf{x}_{i',T^*+1} \right| + \left| \mathbf{x}'_{1,t} \Sigma_{\tilde{\theta}_1} \mathbf{x}_{1,t} - \mathbf{x}'_{i',t} \Sigma_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t} \right| \\ & \quad + \left| \mu_{\tilde{\phi}_1}^2 \left( \sigma_\eta^2 + \sigma_\alpha^2 + \sigma^2 + \mathbf{x}'_{1,T^*+1} \Sigma_\delta \mathbf{x}_{1,T^*+1} + \mathbf{x}'_{1,t-1} \Sigma_{\tilde{\theta}_1} \mathbf{x}_{1,t-1} \right) \right. \\ & \quad \left. - \mu_{\tilde{\phi}_{i'}}^2 \left( \sigma_\eta^2 + \sigma_\alpha^2 + \sigma^2 + \mathbf{x}'_{i',T^*+1} \Sigma_\delta \mathbf{x}_{i',T^*+1} + \mathbf{x}'_{i',t-1} \Sigma_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t-1} \right) \right| \\ & \quad + \left| \mu_{\tilde{\phi}_1}^4 \text{Var}(y_{1,t-2}) - \mu_{\tilde{\phi}_{i'}}^4 \text{Var}(y_{i',t-2}) \right| \\ & \leq \sum_{j=0}^{t-(T^*+1)} \left[ \left| \mu_{\tilde{\phi}_1}^{2j} \mathbf{x}'_{1,T^*+1} \Sigma_\delta \mathbf{x}_{1,T^*+1} - \mu_{\tilde{\phi}_{i'}}^{2j} \mathbf{x}'_{i',T^*+1} \Sigma_\delta \mathbf{x}_{i',T^*+1} \right| \right. \end{aligned}$$

$$\begin{aligned}
& + \left| \mu_{\tilde{\phi}_1}^{2j} \mathbf{x}'_{1,t-j} \Sigma_{\tilde{\theta}_1} \mathbf{x}_{1,t-j} - \mu_{\tilde{\phi}_{i'}}^{2j} \mathbf{x}'_{i',t-j} \Sigma_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t-j} \right| + \left| \mu_{\tilde{\phi}_1}^{2j} - \mu_{\tilde{\phi}_{i'}}^{2j} \right| (\sigma_\eta^2 + \sigma_\alpha^2 + \sigma^2) \Big] \\
& + \left| \mu_{\tilde{\phi}_1}^{2(t-T^*)} \text{Var}(y_{1,t-T^*}) - \mu_{\tilde{\phi}_{i'}}^{2(t-T^*)} \text{Var}(y_{i',t-T^*}) \right| \\
& \leq \sum_{j=0}^{t-(T^*+1)} \left[ \left| \mu_{\tilde{\phi}_1}^{2j} - \mu_{\tilde{\phi}_{i'}}^{2j} \right| \lambda_{\max} + \left| \mathbf{x}'_{1,T^*+1} \Sigma_{\delta} \mathbf{x}_{1,T^*+1} - \mathbf{x}'_{i',T^*+1} \Sigma_{\delta} \mathbf{x}_{i',T^*+1} \right| \mu_{\tilde{\phi}_1}^{2j} \right. \\
& + \left| \mu_{\tilde{\phi}_1}^{2j} - \mu_{\tilde{\phi}_{i'}}^{2j} \right| \lambda_{\max} + \left| \mathbf{x}'_{1,t-j} \Sigma_{\tilde{\theta}_1} \mathbf{x}_{1,t-j} - \mathbf{x}'_{i',t-j} \Sigma_{\tilde{\theta}_{i'}} \mathbf{x}_{i',t-j} \right| \mu_{\tilde{\phi}_1}^{2j} + \left| \mu_{\tilde{\phi}_1}^{2j} - \mu_{\tilde{\phi}_{i'}}^{2j} \right| (\sigma_\eta^2 + \sigma_\alpha^2 + \sigma^2) \Big] \\
& + \left| \mu_{\tilde{\phi}_1}^{2(t-T^*)} \sum_{j=0}^{T^*-1} (\sigma_\eta^2 + \sigma^2 + \mathbf{x}'_{1,T^*-j} \Sigma_{\theta} \mathbf{x}_{1,T^*-j}) \mu_{\tilde{\phi}}^{2j} \right. \\
& \quad \left. - \mu_{\tilde{\phi}_{i'}}^{2(t-T^*)} \sum_{j=0}^{T^*-1} (\sigma_\eta^2 + \sigma^2 + \mathbf{x}'_{i',T^*-j} \Sigma_{\theta} \mathbf{x}_{i',T^*-j}) \mu_{\tilde{\phi}}^{2j} \right| \\
& \leq 2\gamma \frac{\lambda_{\max}}{(1 - \mu_{\tilde{\phi}_1}^2)(1 - \mu_{\tilde{\phi}_{i'}}^2)} + 2\gamma \frac{1}{1 - \mu_{\tilde{\phi}_1}^2} + \gamma \frac{\sigma_\eta^2 + \sigma_\alpha^2 + \sigma^2}{(1 - \mu_{\tilde{\phi}_1}^2)(1 - \mu_{\tilde{\phi}_{i'}}^2)} \\
& + (\sigma_\eta^2 + \sigma^2) \left| \mu_{\tilde{\phi}_1}^{2(t-T^*)} - \mu_{\tilde{\phi}_{i'}}^{2(t-T^*)} \right| \sum_{j=0}^{T^*-1} \mu_{\tilde{\phi}}^{2j} \\
& + \sum_{j=0}^{2(t-T^*)} \left| \mu_{\tilde{\phi}_1}^{2(t-T^*)} \mathbf{x}'_{1,T^*-j} \Sigma_{\theta} \mathbf{x}_{1,T^*-j} - \mu_{\tilde{\phi}_{i'}}^{2(t-T^*)} \mathbf{x}'_{i',T^*-j} \Sigma_{\theta} \mathbf{x}_{i',T^*-j} \right| \mu_{\tilde{\phi}}^{2j} \\
& \leq 2\gamma \frac{\lambda_{\max}}{(1 - \mu_{\tilde{\phi}_1}^2)(1 - \mu_{\tilde{\phi}_{i'}}^2)} + 2\gamma \frac{1}{1 - \mu_{\tilde{\phi}_1}^2} + \gamma \frac{\sigma_\eta^2 + \sigma_\alpha^2 + \sigma^2}{(1 - \mu_{\tilde{\phi}_1}^2)(1 - \mu_{\tilde{\phi}_{i'}}^2)} \\
& + \gamma \frac{\sigma_\eta^2 + \sigma^2}{1 - \mu_{\tilde{\phi}}^2} + \gamma \frac{\lambda_{\max}}{1 - \mu_{\tilde{\phi}}^2} + \gamma \frac{1}{1 - \mu_{\tilde{\phi}}^2}
\end{aligned}$$

We can choose  $\gamma$  small enough so that  $|\text{Var}(y_{1,t}) - \text{Var}(y_{i',t})| < \varepsilon$ .  $\square$

**Lemma 3** (Technical Lemma 2). *Let  $a_j, b_j \in \mathbb{R}$ ,  $j = 1, 2$ . Suppose that  $b_j > b > 0$  for  $j = 1, 2$ . Then for any  $\varepsilon > 0$  we have*

$$\left| \frac{a_1}{b_1} - \frac{a_2}{b_2} \right| \leq \varepsilon,$$

*provided that  $|a_1 - a_2|, |b_1 - b_2| < \delta$  for some chosen  $\delta > 0$ .*

*Proof.*

$$\begin{aligned}
\left| \frac{a_1}{b_1} - \frac{a_2}{b_2} \right| & = \left| \frac{a_1 b_2 - a_2 b_1}{b_1 b_2} \right| = \left| \frac{(a_1 - a_2 + a_2)(b_2 - b_1 + b_1) - a_2 b_1}{b_1 b_2} \right| \\
& = \left| \frac{a_1(b_2 - b_1) + a_2 b_1 + (a_1 - a_2)b_1 - a_2 b_1}{b_1 b_2} \right| \leq \frac{2\delta}{b^2} \\
& \leq \frac{|a_1|\delta + |b_1|\delta}{b^2} \leq \varepsilon,
\end{aligned}$$

where  $\delta \leq \varepsilon b^2 / (|a_1| + |b_1|)$ .  $\square$

**This needs some work (ignore for now)** Now consider a case where there exists a  $2 \leq i' < n+1$  where  $\mathbf{x}_{1,t} = \mathbf{x}_{i',t}$ . Now suppose for simplicity that  $T_{i'}^* = T_1^*$  and that  $T_{i'} = T_1$  where  $T_1 > T_1^* + q_1 + q_2 + 2$ . This assumes that we have data for the first time series beyond the shock point. We will show that when

there is a covariate clone of the time series under study, then estimation in this series will approximate estimation in the time series under study provided that the underlying variability is very small relative to the signal. To see this, consider  $t > T_1^*$ , where

$$\begin{aligned}
y_{1,t} &= \eta_1 + \mu_\alpha + \delta_1' \mathbf{x}_{1,T_1^*+1} + \sum_{j=1}^{q_1} (\phi_{1,j} + \tilde{\phi}_{1,j}) y_{1,t-j} + \sum_{j=0}^{q_2-1} (\theta_{1,j+1} + \tilde{\theta}_{1,j+1})' \mathbf{x}_{1,t-j} + \tilde{\varepsilon}_1 + \varepsilon_{1,t}, \\
y_{i',t} &= \eta_{i'} + \mu_\alpha + \delta_{i'}' \mathbf{x}_{1,T_1^*+1} + \sum_{j=1}^{q_1} (\phi_{i',j} + \tilde{\phi}_{i',j}) y_{i',t-j} + \sum_{j=0}^{q_2-1} (\theta_{i',j+1} + \tilde{\theta}_{i',j+1})' \mathbf{x}_{1,t-j} + \tilde{\varepsilon}_{i'} + \varepsilon_{i',t}, \\
\hat{y}_{1,t} &= \hat{\eta}_1 + \hat{\alpha}_1(\mathbf{x}_{1,T_1^*+1}) + \sum_{j=1}^{q_1} (\hat{\phi}_{1,j} + \hat{\tilde{\phi}}_{1,j}) \hat{y}_{1,t-j} + \sum_{j=0}^{q_2-1} (\hat{\theta}_{1,j+1} + \hat{\tilde{\theta}}_{1,j+1})' \hat{\mathbf{x}}_{1,t-j}, \\
\hat{y}_{i',t} &= \hat{\eta}_{i'} + \hat{\alpha}_{i'}(\mathbf{x}_{1,T_1^*+1}) + \sum_{j=1}^{q_1} (\hat{\phi}_{i',j} + \hat{\tilde{\phi}}_{i',j}) \hat{y}_{i',t-j} + \sum_{j=0}^{q_2-1} (\hat{\theta}_{i',j+1} + \hat{\tilde{\theta}}_{i',j+1})' \hat{\mathbf{x}}_{1,t-j},
\end{aligned}$$

where  $\alpha_j(\mathbf{x}) = \mu_\alpha + \delta_j' \mathbf{x}$ , and we have

$$\begin{aligned}
|y_{1,t} - \hat{y}_{1,t}| &= \left| (\eta_1 - \hat{\eta}_1) + (\alpha_1 - \hat{\alpha}_1(\mathbf{x}_{1,T_1^*+1})) + \sum_{j=1}^{q_1} [(\phi_{1,j} + \tilde{\phi}_{1,j}) y_{1,t-j} - (\hat{\phi}_{1,j} + \hat{\tilde{\phi}}_{1,j}) \hat{y}_{1,t-j}] \right. \\
&\quad \left. + \sum_{j=0}^{q_2-1} [(\theta_{1,j+1} + \tilde{\theta}_{1,j+1})' \mathbf{x}_{1,t-j} - (\hat{\theta}_{1,j+1} + \hat{\tilde{\theta}}_{1,j+1})' \hat{\mathbf{x}}_{1,t-j}] + \varepsilon_{1,t} \right| \\
|y_{i',t} - \hat{y}_{i',t}| &= \left| (\eta_{i'} - \hat{\eta}_{i'}) + (\alpha_{i'} - \hat{\alpha}_{i'}(\mathbf{x}_{1,T_1^*+1})) + \sum_{j=1}^{q_1} [(\phi_{i',j} + \tilde{\phi}_{i',j}) y_{i',t-j} - (\hat{\phi}_{i',j} + \hat{\tilde{\phi}}_{i',j}) \hat{y}_{i',t-j}] \right. \\
&\quad \left. + \sum_{j=0}^{q_2-1} [(\theta_{i',j+1} + \tilde{\theta}_{i',j+1})' \mathbf{x}_{1,t-j} - (\hat{\theta}_{i',j+1} + \hat{\tilde{\theta}}_{i',j+1})' \hat{\mathbf{x}}_{1,t-j}] + \varepsilon_{i',t} \right|.
\end{aligned}$$

**Needs finishing, but the argument is simple: estimation uncertainty declines as error variance decreases. And differences in the random effects decrease as variability decreases.**

### 3 Numerical Examples

In this section, we illustrate shock prevalence testing using two real data examples. Section 3.1 discusses the prevalence testing of the shock that Conoco Phillips stock experienced on March 9th, 2020. Section 3.2 discusses the prevalence testing of the shock that seasonally adjusted log nonfarm payrolls experienced in April, 2020.

#### 3.1 Testing shock prevalence for Conoco Phillips stock price series

Post-shock prevalence testing is motivated by post-shock prediction [Lin and Eck, 2021] and also an extension of it from the perspective of hypothesis testing. Applying our methodology on one of the datasets demonstrated in Lin and Eck [2021] helps connect our work to Lin and Eck [2021] and spell out the differences between two methodologies.

As illustrated in Section 5.1 in Lin and Eck [2021], it is found that Conoco Phillips stock price experienced unprecedented shocks due to oil supply control and COVID-19 pandemic on one single day,



March 9th, 2020. In this data example, we are interested in investigating whether this shock is prevalent, i.e., whether the impact of the shock on the stock price will persist over time. Denote the stock price of Conoco Phillips at time  $t$  as  $y_{i,t}$  for  $i$ th time series, where  $i = 1, \dots, n+1$ , 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.

To analyze the dataset, we propose two models. Denote the covariate vector at time  $t$  for  $i$ th time series by  $\mathbf{x}_{i,t}$ . The first model  $\mathcal{M}_1$ , called “permanent” model, is

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

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

where  $q_1 = q_2 = 2$ ,  $D_{i,t} = I(t > T_i^* + 1)$ ,  $i = 1, \dots, n+1$  and  $t = 1, \dots, T_i$ . Note that we allow the shock effect to take effect for multiple time points instead of one time point as in Lin and Eck [2021]. This simple model describes the setting where one additive shock effect is persistent over time. The second model  $\mathcal{M}_2$ , called “dynamic” model, is

$$y_{i,t} = \left( \eta_i + \sum_{j=1}^{q_1} \phi_{i,j} y_{i,t-j} + \sum_{j=0}^{q_2-1} \mathbf{x}_{i,t-j} \boldsymbol{\theta}_{i,j+1} \right) (1 - D_{i,t}) + f(\mathcal{F}_{i,t}, \alpha_i) D_{i,t} + \varepsilon_{i,t}$$

$$f(\mathcal{F}_{i,t}, \alpha_i) = \alpha_i + \sum_{j=1}^{q_1} \tilde{\phi}_{i,j} y_{i,t-j} + \sum_{j=0}^{q_2-1} \mathbf{x}_{i,t-j} \tilde{\boldsymbol{\theta}}_{i,j+1}$$

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

where  $q_1 = q_2 = 2$ ,  $D_{i,t} = I(t > T_i^* + 1)$ ,  $i = 1, \dots, n+1$ , and  $t = 1, \dots, T_i$ . We estimate  $\mathcal{M}_1$  and  $\mathcal{M}_2$  using ordinary least squares (OLS) estimation.  $\mathcal{M}_2$  adds more flexibility to  $\mathcal{M}_1$  by allowing that the model of  $y_{i,t}$  changes after the shock, as seen from the term related to  $1 - D_{i,t}$ , though the model still preserves the autoregressive model form. In practice,  $\mathcal{M}_2$  should make more sense than  $\mathcal{M}_1$  as it accommodates for more complicated cases. However,  $\mathcal{M}_2$  requires far more degrees of freedom for estimation and this can increase variance in estimation. For this purpose, the synthetic pairwise model selection proposed in the Corollary 2 can be used to select between  $\mathcal{M}_1$  and  $\mathcal{M}_2$ .

In computing the weights, to avoid the impact of unit differences from different columns of covariates on weighting, we center and scale the covariates. The steps are detailed as below. Consider

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_{1,T_1^*+1} & \mathbf{x}_{1,T_1^*} \\ \vdots & \vdots \\ \mathbf{x}_{n+1,T_{n+1}^*+1} & \mathbf{x}_{1,T_1^*} \end{pmatrix},$$

where  $t = T_i^* + 1$  is the time point when the shock effect takes place for  $i$ th time series. Note that the reason for including  $\mathbf{x}_{i,T_i^*}$  is that we consider  $q_2 = 2$  in  $\mathcal{M}_1$  and  $\mathcal{M}_2$ . In other words, we consider the 1-step lagged covariates in the model and its corresponding shock effect time point is  $T_i^*$ . For each column of  $\mathbf{X}$ , we center it by subtracting its column mean and scale the resulting column vector by dividing the standard deviation of the resulting column vector. Denote the resulting covariates matrix by  $\mathbf{X}^*$ . We compute the synthetic weights  $\mathbf{W}^*$  based on  $\mathbf{X}^*$  according to procedures detailed in Section 2.2 of Lin and Eck [2021]. However, for the model of  $y_{i,t}$ , we do not conduct such centering and scaling procedure.

From the definition of  $D_{i,t} = I(t > T_i^* + 1)$ , it is evident that the number of post-shock time points, i.e., time  $t$  for which  $t > T_i^* + 1$ , affects the inference results. As our methodology only requires  $T_i - T_i^*$  is not small compared to  $T_i$ , we attempted three post-shock time length for our inference, say  $L = 7, 15, 30$ . Recall that in Section 2.2, we refer to the forecast that incorporates shock information as adjusted forecast and the one that does not as unadjusted forecast. As specified in Section 2.2, post-shock prevalence testing requires specification of the maximum number  $H$  of forecast horizons and training sample size  $K$ , where  $h$ -step forecasts of adjusted forecast and unadjusted forecast are compared for  $h = 1, \dots, H$ . We set  $H = 10$ ,  $K = 30$ , and  $T_i = H + K + L + 15$  for  $i = 2, \dots, 4$ . That being said, for  $t = K + H + 1, \dots, T_i$  and  $h = 1, \dots, H$ , different  $h$ -step adjusted and unadjusted forecasts for  $y_{i,t}$  under  $\mathcal{M}_1$  or  $\mathcal{M}_2$  are compared.

Table 1: Results for Conoco Phillips ( $n = 3$ ),  $\mathbf{W}^* = (0.183, 0, 0.817)$  from scaled covariates

$L$	Model	$\sum_{i=2}^4 w_i^* I(a_i^1 > a_i^2)$	$p_1$	$ \hat{p} - p_1 $	Vote (Correct?)	weighted Vote (Correct?)
7	$\mathcal{M}_1$	0.817	1	0	0 (Yes)	0 (Yes)
7	$\mathcal{M}_2$		1	0	0 (Yes)	0 (Yes)
15	$\mathcal{M}_1$	0.817	1	0.82	0 (Yes)	1 (No)
15	$\mathcal{M}_2$		1	0	0 (Yes)	0 (Yes)
30	$\mathcal{M}_1$	1	1	0.82	0 (Yes)	1 (No)
30	$\mathcal{M}_2$		1	0	0 (Yes)	0 (Yes)

We implemented our analysis of the Conoco Phillips dataset and summarized the results in Table 1. We discuss our finding in several parts as follows.

Note that the synthetic weight vector  $\mathbf{W}^*$  for the three donors is  $(0.183, 0, 0.817)$ . 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  $\{y_{1,t} : t = 1, \dots, T_1\}$  is due to oil supply shock and recessions caused by COVID-19 pandemic.

In term of Table 1, it is important to note that  $p_1$  is the  $p$ -value of the forecast comparison between adjusted forecast and unadjusted forecast if we have the data  $\{y_{1,t} : 1, \dots, T_1\}$ , where  $T_1 > T_1^*$ . In other words, it cannot be computed without observing  $y_{1,t}$  for  $t > T_1^*$  and we present  $p_1$  only for retrospective analysis. Also,  $\hat{p}$  is the  $p$ -value from weighting the  $p$ -values of the forecast comparison between adjusted and unadjusted forecast for donors using  $\mathbf{W}^*$ .

Let  $a_i^1$  and  $a_i^2$  denote the AICs of  $\mathcal{M}_1$  and  $\mathcal{M}_2$  for  $i$ th time series, respectively. Using the synthetic pairwise model selection procedures specified in Corollary 2, as  $\sum_{i=2}^4 w_i^* I(a_i^1 > a_i^2) > 0.5$  for  $L = 7, 15, 30$ , we conclude that  $\mathcal{M}_2$  is preferred over  $\mathcal{M}_1$ . Notice that  $|\hat{p} - p_1|$  is uniformly zero and  $p_1$  is uniformly 1 for different  $L$  and  $\mathcal{M}_2$ . The test using  $\hat{p}$  with  $\mathcal{M}_2$  fails to reject that the shock, which was to take effect at  $t = T_1^* + 1$ , is prevalent for  $t > T_1^* + 1$  under a significance level of 5%. Nevertheless, the results appear to be a little different for  $\mathcal{M}_1$ . For  $L = 7$ ,  $|\hat{p} - p_1| = 0$  for  $\mathcal{M}_1$  whereas  $|\hat{p} - p_1| = 0.82$  for  $L = 15, 30$ .

Based on the performance of  $\mathcal{M}_2$ , it is possible that as the time window  $L$  increases, the signal from changing parameters covariates is “misunderstood” by  $\mathcal{M}_1$  as shock signal.

Recall from Section 2.2, voting procedure suggests shock transience (shock prevalence) if the majority of the tests suggests shock transience (shock prevalence, respectively). We record 0 if the voting procedure suggests shock transience and 1 if shock prevalence. The weighted voting procedure weight those testing indicators using  $\mathbf{W}^*$ ; and returns shock prevalence if the output exceeds 0.5 and shock transience if the output is smaller than 0.5. As we observe  $\{y_{1,t}: 1, \dots, T_1\}$ , we are able to check whether the decision of voting and weighted voting is correct. From the sixth column of Table 1, we can see that voting procedure suggests shock transience for  $\mathcal{M}_1$  and  $\mathcal{M}_2$  for different  $L$ . However, weighted voting suggests shock transience for  $\mathcal{M}_2$  for different  $L$  but shock prevalence for  $\mathcal{M}_1$  and  $L = 15, 30$  and shock transience for  $\mathcal{M}_1$  and  $L = 7$ . As  $p_1$  is uniformly 1, 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  $n = 3$ . Note that the interpretation and meaning of  $p_1$  is closely connected with models. For example,  $\mathcal{M}_1$  would imply a persistent shock whereas  $\mathcal{M}_2$  is more general in suggesting changing dynamics. Thus, even if this theoretical statement may be correct, it makes practical sense only when model is identified well. This comment corresponds to the possibility that  $\mathcal{M}_1$  may “misunderstand” the wrong signal of changing dynamics that are instead captured by  $\mathcal{M}_2$ .

To verify our results, we plot  $\{y_{1,t}: 1, \dots, T_1\}$  in Figure 1. 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  $t > T_1^*$ . This observation confirms our results as the reverting behavior implies that the impact of the shock starts to diminish as  $t$  increases and that the shock is very likely to be transient.

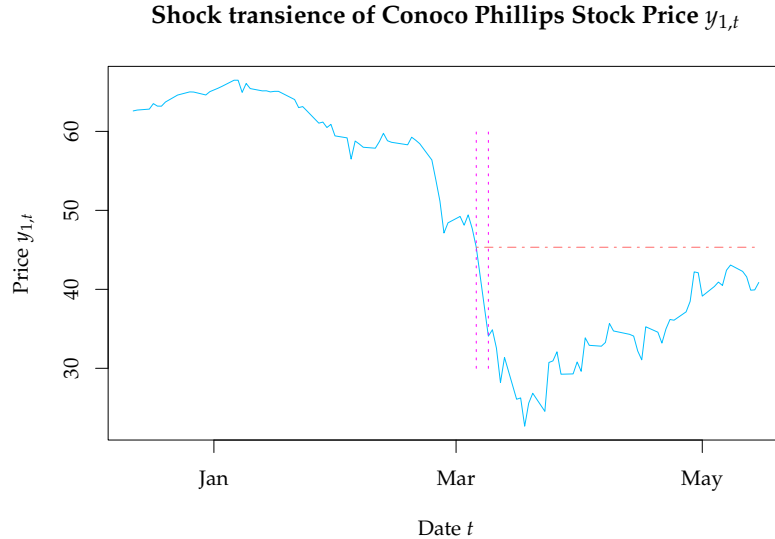


Figure 1: The shock transience of  $\{y_{1,t}: 1, \dots, T_1\}$

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

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. 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 many orders of magnitude, and hence then distance-based weighting employed in [Lin and Eck, 2021] could not hope to recover the full shock effect. However, the test developed herein does not rely as heavily upon the assumption of a common shock distribution for donors or the assumption that the donors represent a rich and diverse sampling from that distribution (**Think about this**). Instead, what is needed is that the donor pool serve as an information source about the persistence of a shock effect in the time series under study.

To that end, we assemble a donor pool of US recessions induced by large shocks in the last three decades: the recessions of 1981-82, 2001-2003, and 2007-2009 (**Justification needed for these shocks; the March 2020 lockdown was an unprecedented halting of economic activity**). We employ a suite of monthly macroeconomic indicators as covariates for the linear model as for arriving 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  $\{\mathcal{X}_1, \dots, \mathcal{X}_p\}$  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  $i = 2, \dots, n+1$ , 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 2, using the synthetic pairwise model selection procedures specified in Corollary 2, because  $\sum_{i=2}^4 w_i^* I(a_i^1 > a_i^2) > 0.5$ , we conclude that  $\mathcal{M}_2$  is preferred. Thus, we turn our interest to that model. Since  $p_1$  is far above the rejection threshold, we lack the evidence to reject the null of a transient shock in the time series under study, the month-on-month change in log employment in the span 2019-2022. The weighted p-value,  $\hat{p}$ , misses the ground truth p-value by .146 in absolute value, yet renders the correct decision.

#### Questions to answer and points to hit

1. Why did I choose the donors I chose? X
2. Why did I use the columns I did in OLS? X
3. Why did I use the columns I did in the estimation of the points on the simplex? X
4. Discuss the two models, M1, M2 X
5. Weighted AIC just as in previous example? X
6. By weighted AIC, M2 is the preferred model Yes
7. We can see that  $p_1$  is far larger than the rejection threshold, indicating that we lack the evidence to reject the null of a transient shock
8. The weighted pvalue,  $\hat{p}$ , misses the ground truth p-value by .146 in absolute value

9. The method of voting renders the correct decision
10. The method of weighted voting renders the correct decision
11. Ask Jilei to correct y-axis X
12. Ask Jilei whether AIC was used in the same way as first example X
13. Ask Jilei whether covariate set  $\mathbf{X}$  is same as simplex optimization  $\mathbf{X}$
14. Ask Jilei to drop RPI as a covariate in OLS (but not necessarily for optimization) (needs checking)
15. Ask Jilei to include chart with data stopping at March 2020 to better visualize how payrolls change by month X

Table 2: Results for Month-on-Month Change in Log Nonfarm Payrolls (Seasonally Adjusted) ( $n = 3$ ),  $\mathbf{W}^* = (0, 0.278, 0.722)$  from scaled covariates

Model	$\sum_{i=2}^4 w_i^* I(a_i^1 > a_i^2)$	$p_1$	$ \hat{p} - p_1 $	Vote (Correct?)	weighted Vote (Correct?)
$\mathcal{M}_1$	0.722	0	1	1 (No)	1 (No)
$\mathcal{M}_2$ ( $q_1 = q_2 = 1$ )		0.470	0.146	0 (Yes)	0 (Yes)

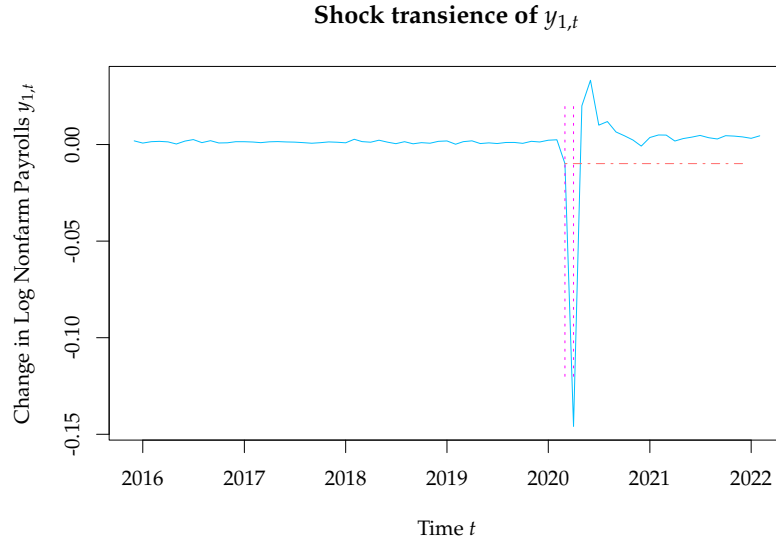


Figure 2: The shock transience of  $\{y_{1,t}: 1, \dots, T_1\}$

### 3.3 Simulations

## 4 Discussion

Talk about post-shock forecasting as parameterizing exogeneity and using hindsight in the donor pool to explicitly estimate exogeneity apriori.

## 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*, 2020.
- 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.
- Raffaella Giacomini and Halbert White. Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578, 2006.
- 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.
- Gary Koop and Dimitris Korobilis. Forecasting inflation using dynamic model averaging. *International Economic Review*, 53(3):867–886, 2012.
- 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.
- 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 Quaadvlieg. 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.
- Paul Shaman and Robert A Stine. The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association*, 83(403):842–848, 1988.
- Lars EO Svensson. Monetary policy with judgment: Forecast targeting. Technical report, National Bureau of Economic Research, 2005.