

Post-shock Volatility Forecasting Using Aggregated Shock Information

David Lundquist*

Department of Statistics, University of Illinois at Urbana-Champaign

December 3rd, 2021

Abstract

We develop a novel procedure for forecasting the volatility of a time series immediately following an exogenous shock event. Adapting the synthetic prediction framework of ?, we exploit series that have experienced similar shock events. We extract volatility signals from these series in order to aggregate properly their shock-induced excess volatilities. The volatility spikes are modeled as random effects and estimated as fixed effects. The aggregation of these estimates is in service of adjusting the 1-step-ahead GARCH forecast of the time series under study by an additive term. The adjusted and unadjusted forecasts are evaluated using three families of benchmarks: the unobservable but easily estimated realized volatility (RV) of the time series under study, implied volatility, and the empirical volatility over horizons of varying length. We also compare the performance of the adjusted forecast to the performance of the Realized GARCH forecast, which is known to react faster to rapidly changing volatility than the standard GARCH model by permitting the inclusion of metrics like implied volatility. Finally, we combine Realized GARCH modeling with the synthetic prediction framework, using Realized GARCH in both the estimation of random effects as well as the forecast for the time series under study. Real world illustrations are provided, as are simulation results suggesting the conditions under which our approach’s hyperparameters can be tuned for best performance.

1 Introduction

The most important stochastic phenomenon of many time series $(X_t)_{t \in \mathbb{N}}$, especially financial time series, is the volatility of the return series of $(X_t)_{t \in \mathbb{N}}$. A financial asset’s price series and daily return series may exhibit behavior that makes inconvenient and uninterpretable the traditional methods of time series analysis. In contrast, the volatility of the return series may be less noisy, less structurally dependent on external factors, and more easily modeled as a weakly stationary process. Yet even if one could model and forecast price series and return series with high accuracy, that would not necessarily tell us much about the variability of such forecasts nor enlighten us about the evolution of the variability of $(X_t)_{t \in \mathbb{N}}$ over time. Modern portfolio management and theory requires information about at least the first two moments of a return series, if not higher. For these reasons and others, volatility modeling has come to dominate financial econometrics over the past four decades.

When $(X_t)_{t \in \mathbb{N}}$ is positive, as one would expect of a price series, the percentage change over any interval of length $k \in \mathbb{N}$ can be approximated by $\log(\frac{X_{t+k}}{X_t})$, owing to the fact that for any nonnegative number y , $\log(y + 1) \leq y$. Equality holds at exactly $y = 0$, corresponding to the case where $X_{t+k} = X_t$. It is easily verified that the approximation works well when the ratio $\frac{X_{t+k}}{X_t} \in (.85, 1.15)$, a return absolute value of 15% or less. The daily log return series $\log(\frac{X_{t+1}}{X_t})$, however, not only provides an approximation to the daily return; the logarithm of a price series is a random variable supported on \mathbb{R} . Partly due to that, it

*davidl11@ilinois.edu

earns a pedigree for its role in geometric Brownian motion, the foundation of much of modern financial mathematics, including the Black-Scholes differential equation (?). Additionally, the log transformation is a tried-and-true variance-reduction technique for heteroskedastic data (?) and non-stationary time series. In many applications, the log-return of a positive time series will allow us to reject the null of a unit-root non-stationary process or trend-stationary process, thus paving the way for the vast family of autoregressive-moving-average models.

However, no matter how a time series or its transformations is modeled, forecasting in the presence of exogenous shock events requires a methodological framework that sensibly incorporates relevant information that has yet to manifest itself in market price or derivative quantities like volatility. In this setting, regime-change models are of little use because under the assumption of a known exogenous shock, there is no need to estimate a regime-change time, nor is there data following the exogenous shock event to fit a model. Asymmetric GARCH models were an early attempt to account for fact that negative returns typically beget larger volatility than positive returns (?). As a separate problem, GARCH models have been shown slow to adapt to spikes in volatility (find first authors to point this out). ? explored the use of implied volatility as an external regressor in a GARCH model, that is, using so-called GARCH-X models. ? propose Realized Garch, which aims to solve both the asymmetry problem as well as the slow-reaction problem by introducing a "measurement equation" as a tool for modeling the contribution to the conditional variance made by the market-implied volatility measure. The key insight is that the market-implied volatility measure is not independent of the conditional variance posited by the GARCH model. Therefore, including the external measure must be done in a way that accounts for this dependence.

The approach herein can be viewed as an attempt to sidestep the functional complexity posited by Realized GARCH by substituting modeling assumptions. Synthetic volatility forecasting proceeds under the assumption that similar exogenous shock events occasion volatility shocks arising from a common shock distribution.

2 Setting

We follow convention and define the daily log-return as $r_t = \log(\frac{X_{t+1}}{X_t})$. The class of ARIMA(p,d,q) models developed in the mid-to-late 20th century provides a framework for postulating and quantifying the autoregressive structure of a r_t , all within the framework of frequentist statistics and maximum-likelihood estimation of parameters. These models assume a certain dependence structure between r_t and $r_{t-\max\{p,q\}}$, as well as all observations in between, yet their errors – often called innovations in the financial time series context due to how they represent the impact of new information – are nevertheless assumed to be i.i.d. with mean zero and constant variance. The ARCH (?) and GARCH (?) models provide elegant alternatives to the constant-variance assumption. In fact, the GARCH framework in its most basic form disregards r_t and instead turns its interest to the series r_t^2 (when properly centered).

To that end, let $a_t = r_t - \mu_t$, where μ_t is the mean of the log return series r_t . Implicitly, we are committing ourselves to a mean-model for r_t in which μ_t may vary with time, while a_t is simply the noise. We derive a mean-zero process a_t with the property that $\mathbb{E}[a_t^2] = \text{Var}[a_t]$. Under the assumption of time-invariant volatility, the series a_t^2 should exhibit no autocorrelation at any lag $\ell \geq 1$. This assumption motivates tests for so-called ARCH effects, that is, tests for the clustering of volatility. These tests explore the alternative hypothesis that σ_t^2 is not only a time-varying parameter but furthermore a function of past squared residuals of the mean model. In particular, the ARCH(m) model is an autoregressive model in which σ_t^2 is fitted using a linear combination of the m past values of a_t^2 . The GARCH(m,s) framework take this one step further but modeling σ_t^2 as a linear combination of the m past a_t^2 values and well as the s past values of σ_t^2 . In functional form,

$$\sigma_t^2 = \omega + \sum_{k=1}^m \alpha_k a_{t-k}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2, a_t = \sigma_t \epsilon_t$$

Assuming further that σ_t^2 depends on a vector of exogenous covariates \mathbf{x}_t , we have

$$\sigma_t^2 = \omega + \sum_{k=1}^m \alpha_k a_{t-k}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 + \gamma^T \mathbf{x}_t, a_t = \sigma_t \epsilon_t$$

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 $\mathbf{y}_{i,t} = (y_{i,t}, \mathbf{x}_{i,t}, \mathbf{v}_{i,t})$ where $y_{i,t}$ is a scalar response, $\mathbf{x}_{i,t}$ is a vector of covariates that are revealed to the analyst prior to the observation of $y_{1,t}$, and $\mathbf{v}_{i,t}$ is a vector of market-implied volatility metrics for the series i that are available prior to the market open at time t . Suppose that the analyst is interested in forecasting the volatility of $y_{1,t}$, the first time series in the collection. We require that each time series $\mathbf{y}_{i,t}$ is subject to a shock event in the time between market close on $T_i^* \leq T_i + 1$ and before the market open on $T_i^* + 1$. To define an interesting setting, we will suppose that $T_1^* = T_1 + 1$, and $1 < T_i^* < T_i + 1$ for $i \geq 2$. We will suppose that $\mathbf{x}_{i,t=T_i^*}$ and $\mathbf{v}_{i,t=T_i^*}$ are observed before the shock is manifested in market activity on $y_{i,t=T_i^*}$.

In light of the foregoing, we can rewrite the GARCH framework we are interested in as such

$$\sigma_{i,t}^2 = \omega_i + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \begin{bmatrix} \mathbf{x}_{i,t} \\ \mathbf{v}_{i,t} \end{bmatrix}, a_{i,t} = \sigma_{i,t} \epsilon_{i,t}$$

We are interested in a point forecast for σ_{1,T^*+1}^2 , the 1-step ahead conditional variance for the time series under study. The forecast performance is evaluated mean squared-error computed using three families of estimators for the ground truth that we seek. Let $L_{\text{predictivemethod,groundtruthestimator}}$ denote the squared-error loss of a predictive method based on a given ground truth estimator. For example, the loss function of primary interest in this study is the 1-step-ahead forecast loss

$$L_{\text{syntheticpredictionGARCH,RV}}^1 = (\hat{\sigma}_{\text{syntheticpredictionGARCH}}^2 - \hat{\sigma}_{\text{RealizedVolatility}}^2)^2$$

In full generality, for an h -step-ahead volatility forecast, the loss function is

$$L_{\text{syntheticpredictionGARCH,RV}}^h = \frac{1}{h} \sum_{g=1}^h (\hat{\sigma}_{g,\text{predictivemethod}}^2 - \hat{\sigma}_{g,\text{groundtruthestimator}}^2)^2$$

2.1 Ground Truth Estimators

2.1.1 Market-Implied Volatility

2.1.2 Realized Volatility

2.1.3 Empirical Volatility

This is just the unbiased estimator of log-return σ^2 over the preceding M days.

2.2 Volatility Profile of a Time Series

In this section we make a novel contribution to the synthetic control framework by constructing a profile of a time series volatility.

2.3 Aggregation Mechanism

Here we explain how we use the volatility profile to arrive at a set of nonnegative weights that sum to 1.

2.4 Model setup

In this section, we will describe the assumed dynamic panel models for which post-shock aggregated estimators are provided. The basic structures of these models are the same for all time series in the analysis, the differences between them lie in the setup of the shock effect distribution.

Let $I(\cdot)$ be an indicator function, T_i be the time length of the time series i for $i = 1, \dots, n+1$, and T_i^* be the time point just before the one when the shock is known to occur, with $T_i^* < T_i$. For $t = 1, \dots, T_i$ and $i = 1, \dots, n+1$, the model \mathcal{M}_1 is defined as

$$\mathcal{M}_1: \sigma_{i,t}^2 = \omega_i + \omega_i^* D_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \begin{bmatrix} \mathbf{x}_{i,t} \\ \mathbf{v}_{i,t} \end{bmatrix}, a_{i,t} = \sigma_{i,t} \epsilon_{i,t} \quad (1)$$

where $D_{i,t} = I(t = T_i^* + 1)$ and $\mathbf{x}_{i,t} \in \mathbb{R}^p$, $\mathbf{v}_{i,t} \in \mathbb{R}^l$ with $p, l \geq 1$. We assume that the $\mathbf{x}_{i,t}$'s are fixed. Let $|x|$ denote the absolute value of x for $x \in \mathbb{R}$. For $i = 1, \dots, n+1$ and $t = 1, \dots, T_i$, the random effects structure for \mathcal{M}_1 is:

$$\begin{aligned} \phi_i &\stackrel{iid}{\sim} \mathcal{F}_\phi \text{ where } |\mathcal{F}_\phi| < 1, [\phi \text{ needs to be replaced with conditions on } \alpha, \beta \text{ which will be simulated fixed parameters}] \\ \gamma_i &\stackrel{iid}{\sim} \mathcal{F}_\gamma \text{ with } E_{\mathcal{F}_\gamma}(\gamma_i) = \mu_\gamma, \text{Var}_{\mathcal{F}_\gamma}(\gamma_i) = \Sigma_\gamma^2 \\ \omega_i^* &\stackrel{iid}{\sim} \mathcal{F}_{\omega^*} \text{ with } E_{\mathcal{F}_{\omega^*}}(\omega_i^*) = \mu_{\omega^*}, \text{Var}_{\mathcal{F}_{\omega^*}}(\omega_i^*) = \sigma_{\omega^*}^2 \\ \varepsilon_{i,t} &\stackrel{iid}{\sim} \mathcal{F}_\varepsilon \text{ with } E_{\mathcal{F}_\varepsilon}(\varepsilon_{i,t}) = 0, \text{Var}_{\mathcal{F}_\varepsilon}(\varepsilon_{i,t}) = 1, \\ \omega_i^* &\perp\!\!\!\perp \phi_i \perp\!\!\!\perp \gamma_i \perp\!\!\!\perp \varepsilon_{i,t}. \end{aligned}$$

Notice that \mathcal{M}_1 assumes that α_i are iid with $E\alpha_i = \mu_\alpha$ for $i = 1, \dots, n+1$. We also consider a model where the shock effects are linear functions of covariates with an additional additive mean-zero error. For $i = 1, \dots, n+1$, the random effects structure for this model (model \mathcal{M}_2) is:

$$\begin{aligned} \mathcal{M}_2: \quad y_{i,t} &= \eta_i + \alpha_i D_{i,t} + \phi_i y_{i,t-1} + \theta_i' \mathbf{x}_{i,t} + \varepsilon_{i,t} \\ \alpha_i &= \mu_\alpha + \delta_i' \mathbf{x}_{i,T_i^*+1} + \tilde{\varepsilon}_i, \end{aligned} \quad (2)$$

where the added random effects are

$$\begin{aligned} \tilde{\varepsilon}_i &\stackrel{iid}{\sim} \mathcal{F}_{\tilde{\varepsilon}} \text{ with } E_{\mathcal{F}_{\tilde{\varepsilon}}}(\tilde{\varepsilon}_i) = 0, \text{Var}_{\mathcal{F}_{\tilde{\varepsilon}}}(\tilde{\varepsilon}_i) = \sigma_\alpha^2 \text{ where } \sigma_\alpha > 0 \\ \eta_i &\perp\!\!\!\perp \alpha_i \perp\!\!\!\perp \phi_i \perp\!\!\!\perp \theta_i \perp\!\!\!\perp \varepsilon_{i,t} \perp\!\!\!\perp \tilde{\varepsilon}_i. \end{aligned}$$

We further define $\tilde{\alpha}_i = \mu_\alpha + \delta_i' \mathbf{x}_{i,T_i^*+1}$. We will investigate the post-shock aggregated estimators in \mathcal{M}_2 in settings where δ_i is either fixed or random. We let \mathcal{M}_{21} denote model \mathcal{M}_2 with $\delta_i = \delta$ for $i = 1, \dots, n+1$, where δ is a fixed unknown parameter. We let \mathcal{M}_{22} denote model \mathcal{M}_2 with the following random effects structure for δ_i :

$$\delta_i \stackrel{iid}{\sim} \mathcal{F}_\delta \text{ with } E_{\mathcal{F}_\delta}(\delta_i) = \mu_\delta, \text{Var}_{\mathcal{F}_\delta}(\delta_i) = \Sigma_\delta \quad \text{with} \quad \delta_i \perp\!\!\!\perp \tilde{\varepsilon}_i.$$

We further define the parameter sets

$$\begin{aligned} \Theta &= \{(\eta_i, \phi_i, \theta_i, \alpha_i, \mathbf{x}_{i,t}, y_{i,t-1}, \delta_i): t = 1, \dots, T_i, i = 2, \dots, n+1\}, \\ \Theta_1 &= \{(\eta_i, \phi_i, \theta_i, \alpha_i, \mathbf{x}_{i,t}, y_{i,t-1}, \delta_i): t = 1, \dots, T_i, i = 1\}, \end{aligned} \quad (3)$$

where Θ and Θ_1 can adapt to \mathcal{M}_1 by dropping δ_i . We assume this for notational simplicity.

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

$$\begin{aligned} \mathcal{M}: \quad y_{i,t} &= \left(\eta_i + \sum_{j=1}^{q_1} \phi_{i,j} y_{i,t-j} + \sum_{j=0}^{q_2-1} \theta'_{i,j+1} \mathbf{x}_{i,t-j} \right) (1 - D_{i,t}) + f(\mathcal{F}_{i,t}, \alpha_i) D_{i,t} + \varepsilon_{i,t}, \\ f(\mathcal{F}_{i,t}, \alpha_i) &= \alpha_i + \sum_{j=1}^{q_1} \tilde{\phi}_{i,j} y_{i,t-j} + \sum_{j=0}^{q_2-1} \tilde{\theta}'_{i,j+1} \mathbf{x}_{i,t-j}, \\ \alpha_i &= \mu_\alpha + \delta_i' \mathbf{x}_{i,T_i^*+1} + \varepsilon_{\alpha,i}, \end{aligned} \quad (4)$$

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 (??) have the following hierarchical random effects structure:

$$\begin{aligned}
\eta_i &\stackrel{iid}{\sim} \mathcal{F}_\eta \text{ with } E_{\mathcal{F}_\eta}(\eta_i) = \mu_\eta, \text{Var}_{\mathcal{F}_\eta}(\eta_i) = \sigma_\eta^2, \\
\phi_{i,j} &\stackrel{iid}{\sim} \mathcal{F}_{\phi_j} \text{ where } |\phi_{i,j}| < 1, \\
\theta_{i,j} &\stackrel{iid}{\sim} \mathcal{F}_{\theta_j} \text{ with } E_{\mathcal{F}_{\theta_j}}(\theta_{i,j}) = \mu_{\theta_j}, \text{Var}_{\mathcal{F}_{\theta_j}}(\theta_{i,j}) = \Sigma_{\theta_j}^2, \\
\delta_i &\stackrel{iid}{\sim} \mathcal{F}_\delta \text{ with } E_{\mathcal{F}_\delta}(\delta_i) = \mu_\delta, \text{Var}_{\mathcal{F}_\delta}(\delta_i) = \Sigma_\delta, \\
\varepsilon_{i,t} &\stackrel{iid}{\sim} \mathcal{F}_{\varepsilon_i} \text{ with } E_{\mathcal{F}_{\varepsilon_i}}(\varepsilon_{i,t}) = 0, \text{Var}_{\mathcal{F}_{\varepsilon_i}}(\varepsilon_{i,t}) = \sigma_i^2, \\
\varepsilon_{\alpha,i} &\stackrel{iid}{\sim} \mathcal{F}_{\varepsilon_\alpha} \text{ with } E_{\mathcal{F}_{\varepsilon_\alpha}}(\varepsilon_{\alpha,i}) = 0, \text{Var}_{\mathcal{F}_{\varepsilon_\alpha}}(\varepsilon_{\alpha,i}) = \sigma_\alpha^2, \\
\tilde{\phi}_{i,j} &\stackrel{ind}{\sim} \mathcal{F}_{\tilde{\phi}_j}(\mathbf{x}_{i,T_i^*+1}) \text{ where } |\tilde{\phi}_{i,j}| < 1, \\
\tilde{\theta}_{i,j} &\stackrel{ind}{\sim} \mathcal{F}_{\tilde{\theta}_j}(\mathbf{x}_{i,T_i^*+1}), \\
\eta_i &\perp\!\!\!\perp \phi_{i,j} \perp\!\!\!\perp \theta_{i,j} \perp\!\!\!\perp \tilde{\phi}_{i,j} \perp\!\!\!\perp \lambda_{i,j} \perp\!\!\!\perp \delta_i \perp\!\!\!\perp \varepsilon_{i,t} \perp\!\!\!\perp \tilde{\varepsilon}_i,
\end{aligned} \tag{5}$$

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 (??) to be of use for post-shock forecasting, the variation in $\mathcal{F}_{\tilde{\phi}_j}$ and $\mathcal{F}_{\tilde{\theta}_j}$ must be small relative to the signal strength.

We see that model (??) with its accompanying random effects structure (??) is flexible enough to capture changing structural dynamics as well as a mean-shift, and is a generalization of the model in ?. 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 $\|\mathbf{x}_{i,T_i^*+1} - \mathbf{x}_{j,T_i^*+1}\|_2$ are expected to experience similar structural changes. This makes distance based weighting an attractive avenue.

2.5 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 ? to conduct shock persistence testing.

In this paper, we extend the framework of post-shock forecast (?), 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 ?. 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 ?. 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 ? 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 refers 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 ?. As introduced in the beginning of Section ??, 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, ? 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 (?). 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 ? 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 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 $\mathbb{P}(p_1 \leq \alpha) \neq 0.5$,

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

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*

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

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 $i \in \mathcal{I}_n$, 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 $\mathbb{P}(p_1 \leq \alpha) \neq 0.5$,

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

Proof. The proof is rather similar to Proposition ???. 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 ?? 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 ?? 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 ?? . 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. □

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 (??) and (??) is large relative to variation. We first present some intermediary technical results and definitions. Let λ_{\max,θ_k} , $\lambda_{\max,\theta_\delta}$, $\lambda_{\max,\gamma}$ respectively be the largest eigenvalues of Σ_{θ_k} , Σ_δ , and Σ_γ for $k = 1, \dots, q_2$. Define

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

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 (??) with accompanying random effects structure (??), 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*

$$\mathbb{E}(y_{i,t}) \geq C \text{Var}(y_{i,t}), \quad (6)$$

provided that

$$\begin{aligned}\mu_\alpha + 3\mu_{\min} &\geq C(3\sigma_{\max}^2 + 2\lambda_{\max}), \\ 2\mu_{\min} &\geq C(2\sigma_{\max}^2 + \lambda_{\max}).\end{aligned} \quad (7)$$

Remarks:

1. The conditions (??) 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 ?? 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 (??) 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 (??) with accompanying random effects structure (??), 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

$$\begin{aligned} |\mu'_{\delta} \mathbf{x}_{1,T^*+1} - \mu'_{\delta} \mathbf{x}_{i',T^*+1}| &< \gamma, \\ \left| \mu'_{\theta_1} \mathbf{x}_{1,t} - \mu'_{\theta_{i'}} \mathbf{x}_{i',t} \right| &< \gamma, \quad \text{for all } t, \\ |\mu_{\tilde{\phi}_1}^j - \mu_{\tilde{\phi}_{i'}}^j| &< \gamma, \quad \text{for all } j, \\ |\mathbf{x}'_{1,T^*+1} \Sigma_{\delta} \mathbf{x}_{1,T^*+1} - \mathbf{x}'_{i',T^*+1} \Sigma_{\delta} \mathbf{x}_{i',T^*+1}| &< \gamma, \\ |\mathbf{x}'_{1,t} \Sigma_{\theta_1} \mathbf{x}_{1,t} - \mathbf{x}'_{i',t} \Sigma_{\theta_{i'}} \mathbf{x}_{i',t}| &< \gamma, \quad \text{for all } t. \end{aligned}$$

Proof. A similar recursion to that seen in the proof of Proposition ?? 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'_{\theta_1} \mathbf{x}_{1,t-j}) - \mu_{\tilde{\phi}_{i'}}^j (\mu'_{\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'_{\theta_1} \mathbf{x}_{1,t} - \mu'_{\theta_{i'}} \mathbf{x}_{i',t}| + \sum_{j=1}^{t-(T^*+1)} \left| \mu_{\tilde{\phi}_1}^j (\mu'_{\theta_1} \mathbf{x}_{1,t-j}) - \mu_{\tilde{\phi}_{i'}}^j (\mu'_{\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'_{\theta_1} \mathbf{x}_{1,t} - \mu'_{\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'_{\theta} \mathbf{x}_{1,t-j}, \mu'_{\theta} \mathbf{x}_{i',t-j}) + |\mu'_{\theta_1} \mathbf{x}_{1,t-j} - \mu'_{\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 ?? . 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 ?? . We can choose γ 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. \\
& \quad + \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] \\
& \quad + \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 \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)} \\
& \quad + (\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} \\
& \quad + \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)} \\
& \quad + \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 γ 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

Corollary 1. *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 (??) with accompanying random effects structure (??), 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. Assume that $\text{Var}(y_{1,t}), \text{Var}(y_{i',t}) \geq b$ for all t . Assume the conditions of Proposition ?? . Then we can choose an $\varepsilon > 0$ such that

$$\left| \frac{\mathbb{E}(y_{1,t})}{\text{Var}(y_{1,t})} - \frac{\mathbb{E}(y_{i',t})}{\text{Var}(y_{i',t})} \right| \leq \varepsilon,$$

for all $1 \leq t \leq T$.

Proof. From Proposition ?? we can pick an arbitrary number $\delta > 0$ such that $|\mathbb{E}(y_{i,t} - y_{i',t})| \leq \delta$ and $|\text{Var}(y_{1,t}) - \text{Var}(y_{i',t})| < \delta$ for all $t = 1, \dots, T$. The conclusion follows from Lemma ?? . \square

This is some justification for the idea that one may be able to make inferences about the time series of interest using a time series from the donor pool.

Doodles:

$$\begin{aligned} \mathbb{P}(|y_{1,t} - \mathbb{E}(y_{1,t})| - |y_{i',t} - \mathbb{E}(y_{i',t})| \leq \varepsilon_1) &\geq \mathbb{P}(|y_{1,t} - \mathbb{E}(y_{1,t})| \leq \varepsilon_1/2, |y_{i',t} - \mathbb{E}(y_{i',t})| \leq \varepsilon_1/2) \\ &= \mathbb{P}(|y_{1,t} - \mathbb{E}(y_{1,t})| \leq \varepsilon_1/2) \mathbb{P}(|y_{i',t} - \mathbb{E}(y_{i',t})| \leq \varepsilon_1/2) \\ &\geq \left(1 - \frac{4 \text{Var}(y_{1,t})}{\varepsilon_1^2}\right) \left(1 - \frac{4 \text{Var}(y_{i',t})}{\varepsilon_1^2}\right) \\ &\geq \left(1 - \frac{4\mathbb{E}(y_{1,t})}{\varepsilon_1^2 C}\right) \left(1 - \frac{4\mathbb{E}(y_{i',t})}{\varepsilon_1^2 C}\right), \end{aligned}$$

for all $t = 1, \dots, T$ where the second inequality follows from Chebyshev's inequality and the last inequality follows from Proposition ?? .

The above argument implies that a C can be chosen so that

$$\begin{aligned} |y_{1,t} - y_{i',t}| &= |y_{1,t} - y_{i',t} \pm \mathbb{E}(y_{1,t}) \pm \mathbb{E}(y_{i',t})| \\ &= |(y_{1,t} - \mathbb{E}(y_{1,t})) - (y_{i',t} - \mathbb{E}(y_{i',t})) + (\mathbb{E}(y_{1,t}) - \mathbb{E}(y_{i',t}))| \\ &\leq |y_{1,t} - \mathbb{E}(y_{1,t})| + |y_{i',t} - \mathbb{E}(y_{i',t})| + |\mathbb{E}(y_{1,t}) - \mathbb{E}(y_{i',t})| \\ &\leq \varepsilon_2, \end{aligned}$$

holds with high probability for all $t = 1, \dots, T$ by Proposition ?? .

Add condition that $\arg\max_t \|\mathbf{x}_{1,t} - \mathbf{x}_{i',t}\| < \delta$. Then the OLS estimates satisfy

$$\|(\hat{\eta}_1, \hat{\phi}_1, \hat{\theta}'_1, \hat{\alpha}_1)' - (\hat{\eta}_{i'}, \hat{\phi}_{i'}, \hat{\theta}'_{i'}, \hat{\alpha}_{i'})'\| < \varepsilon_3$$

with high probability, and therefore $|y_{1,t} - \hat{y}_{i,t}| - |y_{i',t} - \hat{y}_{i',t}| < \varepsilon_4$ with high probability for all $t = 1, \dots, T$.

3 Simulation Setup

Let n denote the donor pool size, p denote the number of covariates used, H denote the number of horizon used, T_i denote the length of time series to be evaluated for time series i , K_i denote the training sample size used for each forecasting time series i , T_i^* denote the time point just before the realization of the shock for time series i for $i = 1, \dots, n + 1$.

In this setting n , p , and H are pre-determined. $T_i, K_i \sim \text{Gamma}(15, 10)$. The total sample size for i th time series is $T_i + K_i + H$. T_i^* is randomly sampled from $\lceil \frac{1}{4}T_i \rceil + 1$ to $\lceil \frac{3}{4}T_i \rceil + K_i + H$. If $T_i, K_i < 90$, we force them to be 90. The adopted model for the data is as below:

$$\begin{aligned} y_{i,t} &= \eta_i + \phi_i y_{i,t-1} + \mathbf{x}_{i,t} \boldsymbol{\beta}_i + \alpha_i I(t > T_i^*) + \varepsilon_{i,t}, \\ \alpha_i &= \mu_\alpha + \mathbf{x}_{i,T_i^*+1} \boldsymbol{\gamma}_i + \tilde{\varepsilon}_i, \end{aligned}$$

where

$$\begin{aligned} \phi_i &\sim \text{indep. } U(0, 1) \\ \eta_i &\sim \text{indep. } \mathcal{N}(0, 1) \\ \varepsilon_{i,t} &\sim \text{indep. } \mathcal{N}(0, \sigma^2) \\ \tilde{\varepsilon}_i &\sim \text{indep. } \mathcal{N}(0, \sigma_\alpha^2) \\ \boldsymbol{\gamma}_i &\sim \text{indep. } \mathcal{N}(\mu_\gamma \mathbf{1}_p, \sigma_\gamma^2 \mathbf{I}_p) \\ \boldsymbol{\beta}_i &\sim \text{indep. } \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p). \end{aligned}$$

Moreover, the elements of $\mathbf{x}_{i,t}$ are independently distributed as $\text{Gamma}(1, \delta)$.

Note that K_i is training sample size for time series i . Consider

$$\begin{aligned} K_i &\sim \lceil \text{Gamma}(a_K, b_K) \rceil \\ T_i &\sim \lceil \text{Gamma}(a_T, b_T) \rceil \\ T_i^* &\equiv \max\{T_i + 1, \lceil 0.5 \cdot (T_i + K_i + H) \rceil\}, \end{aligned}$$

$$K_i + H + T_i^* > T_i + K_i + H$$

Then, we consider the following simulation setup

```
ns <- c(5, 10, 20, 40)
Tscale <- Kscale <- 1 / 2 # b_T, b_K
K.T.shape <- c(200, 400, 800, 1600) # for K_i and T_i
mu.gamma.delta <- 2 # mean for parameter vector of shock
sigma.delta.gamma <- 0.1 # sd for parameter vector of shock
sigma.alpha <- 0.05 # sd for shock noise
sigma <- 0.1 # sd for response noise
mu.alpha <- 50 # intercept for shock (relatively large)
H <- 8
ell <- 4
scale <- 2 # scale for covariates that follow Gamma distribution
```

$$\begin{aligned} y_{i,t} &= \eta_i + \phi_i y_{i,t-1} + \mathbf{x}_{i,t} \boldsymbol{\beta}_i + \xi_i \cdot I(t > T_i^*) + \varepsilon_{i,t}, \\ \xi_i &= \alpha_i \cdot e^{-(t-T_i^*-1)} \\ \alpha_i &= \mu_\alpha + \mathbf{x}_{i,T_i^*+1} \boldsymbol{\gamma}_i + \tilde{\varepsilon}_i, \end{aligned}$$