**Note**

Things to do:

1. Literature review and introduction (We need a literature review for existing methods that are similar/different but close enough methods. Possible searches include pooling, time series pooling, bayesian time series, bayesian autoregression)

2. Explore bootstrapping

3. Data analysis

In R, `optim` function in `base` and `Rsolnp` package may be helpful for solving the minimization problems in synthetic control methods.

**Updates in this version**

1. The manuscript is revised according to the comments and unnecessary parts are deleted.

2. Parametric bootstrap for AR(1) is added in Section 3.3 with some references added.

3. A couple of references are added in the introduction but more to be added.

# Contents

# Minimizing post shock forecasting error using disparate information

Jilei Lin[*], Ziyu Liu[†], and Daniel J. Eck[‡]

Department of Statistics, University of Illinois at Urbana-Champaign

May 18, 2020

**Abstract**

We develop a forecasting methodology for time series data that has undergone a shock. We still can provide credible forecasts for a time series in the presence of such systematic shocks by drawing from disparate time series that have undergone similar shocks for which post-shock outcome data is recorded. These disparate time series are assumed to have mechanistic similarities to the time series under study but are otherwise independent (Granger noncausal). The inferential goal of our forecasting methodology is to supplement observed time series data with post-shock data from the disparate time series in order to minimize average forecast risk.

## 1 Introduction

The technique of combining forecasts to lower forecast error has a rich history [Bates and Granger, 1969, Mundlak, 1978, Timmermann, 2006, Granger and Newbold, 2014]. The Introduction of Timmermann [2006] provided several reasons for combining forecasts. In particular, combining forecasts may be beneficial when: 1) the information set underlying individual forecasts is often unobserved to the forecast user; 2) different individual forecasts may be very differently affected by non-stationarities and model misspecifications; 3) different individual forecasts may be motivated by different loss functions [Timmermann, 2006, and references therein]. The setting for the forecast combination problem is that there are competing forecasts for a single time series. In this setting, one may desire combining forecasts as a method for lowering overall forecast error.

In this article we provide forecasting adjustment techniques with the goal of lowering overall forecast error when the time series under study has undergone a structural shock. It is unlikely that any forecast that previously gave successful predictions for the time series of interest will be able to accommodate the structural shock. Therefore the traditional forecast combination framework may not be of any help. However, all is not lost in this setting. It may be the case that there exists disparate time series that have previously undergone similar structural shocks. When this is so, one may be able to aggregate the post-shock information from these disparate time series to aid the post-shock forecast for the time series under investigation.

We develop and compare aggregation techniques in this post-shock setting and investigate settings for when they do and do not decrease mean squared prediction error. We assume a simple auto regressive data generating process with a general random effects structure. The main

[*]jileil2@ilinois.edu

[†]ziyuliu3@illinois.edu

[‡]dje13@illinois.edu

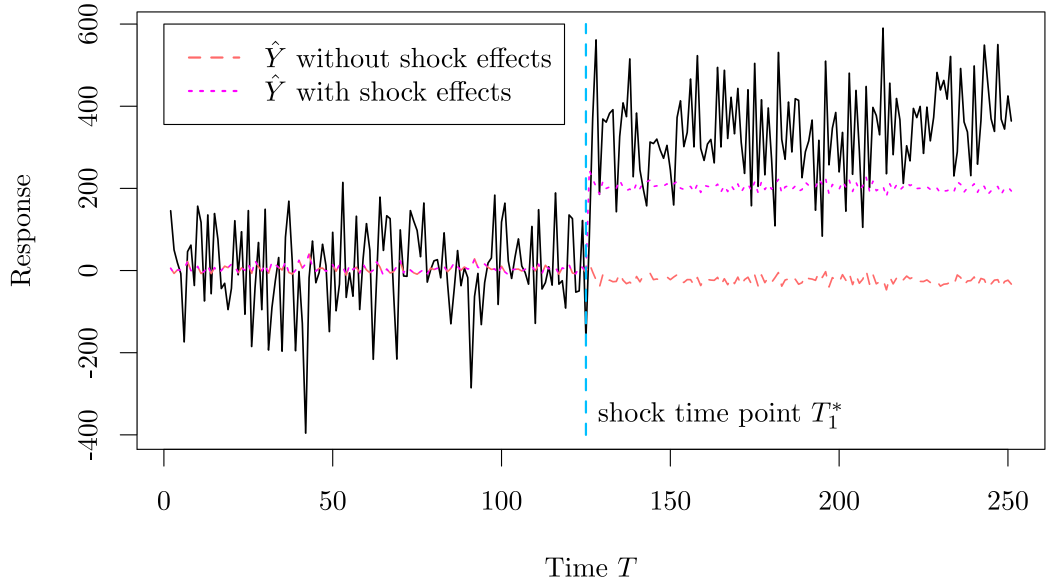


**Figure 1.** A comparison between forecast without considering shock effects and the one uses simple averaging given $n = 40$ disparate time series and that the shock time is at $T_1^* = 125$.

idea is to first average the estimated shock effects from the disparate time series and then add the averaged estimated shock effect to the present forecast. When these time series are independent and the mean of shock effect distribution is large relative to its variance then this technique will reduce mean squared prediction error under the assumed model. Note that this methodology is not motivated with the goal of unbiased, asymptotically unbiased, or consistent estimation for the shock-effect of the time series under study. We consider three aggregation techniques: simple averaging, inverse-variance weighted averaging, and similarity weighting. The latter technique is similar to the weighting in synthetic control methodology [Abadie et al., 2010].

*More to be added ...*

Time-series pooling methodology is one of the main methods to deal this problem. The literature of time-series pooling is mainly related to pooling cross-sectional data. In this setting, estimation of parameters is usually based on merging from different time series using various pooling techniques. For example, Zellner et al. [1991] assumes the coefficient vectors of different time series are the same so that the time series can be merged to estimate parameters once in forecasting turning point of GDP growth rate in practice. *More to be added ...*

Modern research in "combining" forecasts from disparate time series have taken perspectives from distinct fields. Lee et al. [2020] constructed a Bayesian hierarchical model to estimate posterior parameters for Richards model to improve predictive precision of COVID-19 infection trajectories for different countries. Though a Bayesian hierarchical approach can be a solution for our problem by setting up prior for shock-effects, the analysis is sensitive to the prior and the hierarchical model setup, and conditions for predictive improvement are generally hard to know. Plessen [2020] employs a data-mining approach to take COVID-19 data from different countries as input to predict *global* net daily infections and deaths of COVID-19 from clustering. However, the fit is poor due to tremendous volatility of COVID-19 data, and perhaps a lack of random structure in his model. *More to be added ...*

## 2 Setting

We will suppose that an analyst has time series data $(y_{i,t}, \mathbf{x}_{i,t})$, $t = 1, \ldots, T_i$, $i = 1, \ldots, n+1$, 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 $y_{1,t}$. Suppose that the analyst is interested in forecasting $y_{1,t}$, the first time series in the collection. To gauge the performance of a procedure that produces forecasts $\{\hat{y}_{1,t}, t = 1, 2, \ldots\}$ given time horizon $T_1$, we consider the average forecast risk

$$R_T = \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}(\hat{y}_{1,t} - y_{1,t})^2$$

in our analyses. In this article, we consider a similar dynamic panel data model with autoregressive structure to that in Blundell and Bond [1998]. Our dynamic panel model includes an additional shock effect whose presence or absence is given by the binary variable $D_{i,t}$, the details of this model are in the next section.

Figure 1 provides simple intuition of the practical usefulness of our proposed methodology. This figure depicts a time-series that experienced a "shock" at time point $T_1^* = 125$. It is supposed that the researcher does not have any information beyond $T_1^*$, but does have observations of forty disparate time series that have previously undergone a similar shock for which post-shock responses are recorded. Similarity in this context means that the shock effects are random variables that from a common distribution. In this example, the mean of the estimated shock effects is taken as a shock-effect estimator for the time series under study. Forecasts are then made by adding this shock-effect estimator to the estimated response values obtained from the process that ignores the shock. It is apparent from Figure 1 that adjusting forecasts in this manner 1) leads to a reduction in forecasting risk; 2) does not fully recover the true shock-effect. We evaluate the performance of this post-shock prediction methodology throughout this article; we outline situations for when it is expected to work and when it is not.

### 2.1 Model Setup

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

The model $\mathcal{M}_1$ is defined as

$$\mathcal{M}_1 \colon y_{i,t} = \eta_i + \alpha_i D_{i,t} + \phi_i y_{i,t-1} + \theta_i' \mathbf{x}_{i,t} + \beta_i' \mathbf{x}_{i,t-1} + \varepsilon_{i,t} \tag{1}$$

for $t = 1, \ldots, T_i$ and $i = 1, \ldots, n+1$, where $D_{i,t} = 1(t > T_i^*)$, $T_i^* < T_i$ and $\mathbf{x}_{i,t} \in \mathbb{R}^p$, $p \geq 1$. We assume that the $\mathbf{x}_{i,t}$'s are fixed and $T_i^*$s are known. The random effects structure for $\mathcal{M}_1$ is:

$$\begin{aligned}
\eta_i &\overset{iid}{\sim} \eta, \text{ where } \mathrm{E}(\eta) = 0, \mathrm{Var}(\eta) = \sigma_\eta^2, \qquad i = 1, \ldots, n+1,\\
\phi_i &\overset{iid}{\sim} \phi, \text{ where } |\phi| < 1, \qquad i = 1, \ldots, n+1,\\
\theta_i &\overset{iid}{\sim} \theta, \text{ where } \mathrm{E}(\theta) = \mu_\theta, \mathrm{Var}(\theta) = \Sigma_\theta^2, \qquad i = 1, \ldots, n+1,\\
\beta_i &\overset{iid}{\sim} \beta, \text{ where } \mathrm{E}(\beta) = \mu_\beta, \mathrm{Var}(\beta) = \Sigma_\beta^2, \qquad i = 1, \ldots, n+1,\\
\varepsilon_{i,t} &\overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \qquad t = 1, \ldots, T_i,\ i = 1, \ldots, n+1,\\
\alpha_i &\overset{iid}{\sim} \mathcal{N}(\mu_\alpha, \sigma_\alpha^2), \qquad i = 1, \ldots, n+1;
\end{aligned}$$

$$\eta \perp\!\!\!\perp \alpha_i \perp\!\!\!\perp \phi \perp\!\!\!\perp \theta \perp\!\!\!\perp \varepsilon_{i,t}.$$

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

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

for $i = 1, \ldots, n+1$, where the added random effects are

$$\begin{aligned} &\tilde{\varepsilon}_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\alpha^2), \qquad i = 1, \ldots, n+1; \\ &\eta \perp\!\!\!\perp \alpha_i \perp\!\!\!\perp \phi \perp\!\!\!\perp \theta \perp\!\!\!\perp \varepsilon_{i,t} \perp\!\!\!\perp \tilde{\varepsilon}_i. \end{aligned}$$

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

$$\begin{aligned} \gamma_i &\overset{iid}{\sim} \mathrm{E}(\gamma) = \mu_\gamma, \mathrm{Var}(\gamma) = \Sigma_\gamma \\ \delta_i &\overset{iid}{\sim} \mathrm{E}(\delta) = \mu_\delta, \mathrm{Var}(\delta) = \Sigma_\delta \end{aligned} \qquad \text{with} \quad \delta_i \perp\!\!\!\perp \tilde{\varepsilon}_i \quad \text{and} \quad \gamma_i \perp\!\!\!\perp \tilde{\varepsilon}_i.$$

Note that $\delta_i$ and $\gamma_i$ may be dependent. We further define the parameter sets

$$\begin{aligned} \Theta &= \{(\eta_i, \phi_i, \theta_i, \beta_i, \alpha_i, \mathbf{x}_{i,t}, y_{i,t-1}, \delta_i, \gamma_i)\colon t = 1, \ldots, T_i, i = 2, \ldots, n+1\}. \\ \Theta_1 &= \{(\eta_i, \phi_i, \theta_i, \beta_i, \alpha_i, \mathbf{x}_{i,t}, y_{i,t-1}, \delta_i, \gamma_i)\colon t = 1, \ldots, T_i, i = 1\}., \end{aligned} \tag{3}$$

where $\Theta$ and $\Theta_1$ can adapt to $\mathcal{M}_1$ by dropping $\delta_i$ and $\gamma_i$. We assume this for notational simplicity.

## 2.2 Forecast

In this section we show how post-shock aggregate estimators improve upon standard forecasts that do not account for the shock effect. More formally, we will consider the following candidate forecasts:

$$\begin{aligned} &\text{Forecast 1}: \hat{y}^1_{1,T_1^*+1} = \hat{\eta}_1 + \hat{\phi}_1 y_{1,T_1^*} + \hat{\theta}_1' \mathbf{x}_{1,T_1^*+1} + \hat{\beta}_1' \mathbf{x}_{1,T_1^*}, \\ &\text{Forecast 2}: \hat{y}^2_{1,T_1^*+1} = \hat{\eta}_1 + \hat{\phi}_1 y_{1,T_1^*} + \hat{\theta}_1' \mathbf{x}_{1,T_1^*+1} + \hat{\beta}_1' \mathbf{x}_{1,T_1^*} + \hat{\alpha}, \end{aligned}$$

where $\hat{\eta}_1$, $\hat{\phi}_1$, $\hat{\theta}_1$, and $\hat{\beta}_1$ are all OLS estimators of $\eta_1$, $\phi_1$, $\theta_1$, and $\beta_1$ respectively, and $\hat{\alpha}$ is some form of estimator for the shock effect of time series of interest, i.e., $\alpha_1$. The first forecast ignores the presence of $\alpha_1$ while the second forecast incorporates an estimate of $\alpha_1$ that is obtained from the other individual forecasts under study.

Note that the two forecasts do not differ in their predictions for $y_{1,t}$, $t = 1, \ldots T_1^*$, they only differ in predicting $y_{1,T_1^*+1}$. Throughout the rest of this article we show that the collection of disparate time series $\{y_{i,t}, t = 2, \ldots, T_i, i = 1, \ldots, n\}$ has the potential to improve the forecasts for $y_{1,t}$ when $t > T_1^*$ under different circumstances for the dynamic panel model $\mathcal{M}_1$, $\mathcal{M}_{21}$, and $\mathcal{M}_{22}$. It is important to note that in general $\hat{\alpha}$ is not a consistent estimator of the unobserved $\alpha_1$ nor does it converge to $\alpha_1$. Despite these inferential shortcomings, adjustment of the forecast for $y_{1,T_1^*+1}$ through the addition of $\hat{\alpha}$ has the potential to lower forecast risk under several conditions corresponding to different estimators of $\alpha_1$.

### 2.3 Construction of shock effects estimators

We now construct the aggregate estimators of the shock effects that appear in Forecast 2. We use these to forecast response values $y_{1,t}$ when $t > T_1^*$, i.e., the time series of interest after the shock time where we assume that $T_1^*$ is known. First, we introduce the procedures of parameter estimation for $\mathcal{M}_1$, $\mathcal{M}_{21}$, and $\mathcal{M}_{22}$ (see Section 2.1). Conditional on all regression parameters, previous responses, and covariates, the response variable $y_{i,t}$ in $\mathcal{M}_1$, $\mathcal{M}_{21}$, and $\mathcal{M}_{22}$ has distribution

$$y_{i,t} \sim N(\eta_i + \alpha_i D_{i,t} + \phi_i y_{i,t-1} + \theta_i' \mathbf{x}_{i,t} + \beta_i' \mathbf{x}_{i,t-1}, \sigma^2).$$

For $i = 2, \ldots, n$, all parameters in this model will be estimated with ordinary least squares (OLS) using historical data of $t = 1, \ldots, n_i$. For $i = 1$, we estimate all the parameters but $\alpha_1$ using OLS procedures for $t = 1, \ldots, T_1^*$. In particular, let $\hat{\alpha}_i$, $i = 2, \ldots, n+1$ be the OLS estimate of $\alpha_i$. Note that parameter estimation for $\mathcal{M}_1$ is identically the same as $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$.

Second, we introduce the candidate estimators for $\alpha_1$. Define the *adjustment estimator* for time series $i = 1$ by,

$$\hat{\alpha}_{\text{adj}} = \frac{1}{n} \sum_{i=2}^{n+1} \hat{\alpha}_i, \tag{4}$$

where the $\hat{\alpha}_i$s in (4) are OLS estimators of all of the $\alpha_i$s. We can use $\hat{\alpha}_{\text{adj}}$ as an estimator for the unknown $\alpha_1$ term for which no meaningful estimation information otherwise exists. It is intuitive that $\hat{\alpha}_{\text{adj}}$ should perform well under $\mathcal{M}_1$ where we assume that $\alpha_i$'s share the same mean for $i = 1, \ldots, n+1$. However, it can also be shown that $\hat{\alpha}_{\text{adj}}$ may be less favorable in $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$, which will be discussed in detail in Section 3.

We also consider the *inverse-variance weighted estimator* in practical settings where the $T_i$'s and $T_i^*$'s vary greatly across $i$. The inverse-variance weighted estimator is defined as

$$\hat{\alpha}_{\text{IVW}} = \frac{\sum_{i=2}^{n+1} \hat{\alpha}_i / \hat{\sigma}_{i\alpha}^2}{\sum_{i=2}^{n+1} 1/\hat{\sigma}_{i\alpha}^2}, \qquad \text{where} \qquad \hat{\sigma}_{i\alpha}^2 = \hat{\sigma}_i^2 (\mathbf{U}_i' \mathbf{U}_i)_{22}^{-1},$$

$\hat{\alpha}_i$ is the OLS estimator of $\alpha_i$, $\hat{\sigma}_i$ is the residual standard error from OLS estimation, and $\mathbf{U}_i$ is the design matrix for OLS with respect to time series for $i = 2, \ldots, n+1$. Note that since $\sigma$ is unknown, estimation is required and the numerator and denominator terms are dependent in general. However, $\hat{\alpha}_{\text{IVW}}$ can be a reasonable estimator in practical settings. We do not provide closed form expressions for $\text{E}(\hat{\alpha}_{\text{IVW}})$ and $\text{Var}(\hat{\alpha}_{\text{IVW}})$, empirical performance of $\hat{\alpha}_{\text{IVW}}$ is assessed via Monte Carlo simulation (see Section 4).

We now motivate a *weighted-adjustment estimator* for model $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$. Our weighted-adjustment estimator is inspired by the weighting techniques in synthetic control methodology (SCM) developed in Abadie et al. [2010]. However, our weighted-adjustment estimator is not a causal estimator and our estimation premise is a reversal of that in SCM. Our objective is in predicting a post-shock response $y_{1,T_1^*+1}$ that is not yet observed using disparate time series whose post-shock responses are observed.

We use similar notation as that in Abadie et al. [2010] to motivate our weighted-adjustment estimator. Consider a $n \times 1$ weight vector $\mathbf{W} = (w_2, \ldots, w_{n+1})'$, where $w_i \in [0, 1]$ for all $i = 2, \ldots, n+1$. Construct

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{x}_{1,T_1^*-1} \\ \mathbf{x}_{1,T_1^*} \end{pmatrix} \qquad \text{and} \qquad \mathbf{X}_0 = \begin{pmatrix} \mathbf{x}_{2,T_2^*-1} & \cdots & \mathbf{x}_{n+1,T_{n+1}^*-1} \\ \mathbf{x}_{2,T_2^*} & \cdots & \mathbf{x}_{n+1,T_{n+1}^*} \end{pmatrix},$$

where $\mathbf{X}_1$ is $2 \times 1$ and $\mathbf{X}_0$ is $2 \times n$. Define $\mathcal{W} = \{\mathbf{W} \in [0,1]^n \colon w_2 + \cdots + w_{n+1} = 1\}$. Suppose there exists $\mathbf{W}^* \in \mathcal{W}$ with $\mathbf{W}^* = (w_2^*, \ldots, w_{n+1}^*)'$ such that

$$\mathbf{X}_1 = \mathbf{X}_0 \mathbf{W} \quad i.e., \quad \mathbf{x}_{1,T_1^*-1} = \sum_{i=2}^{n+1} w_i^* \mathbf{x}_{i,T_i^*-1} \text{ and } \mathbf{x}_{1,T_1^*} = \sum_{i=2}^{n+1} w_i^* \mathbf{x}_{i,T_i^*}. \tag{5}$$

Notice that $\mathbf{W}^*$ exists as long as $\mathbf{X}_1$ falls in the convex hull of

$$(\mathbf{x}_{2,T_2^*-1} \cdots \mathbf{x}_{n+1,T_{n+1}^*-1}, \ldots, \mathbf{x}_{2,T_2^*}, \cdots \mathbf{x}_{n+1,T_{n+1}^*}). \tag{6}$$

Our weighted-adjustment estimator will therefore perform well when the pool of disparate time series posses similar covariates to the time series for which no post-shock responses are observed. We compute $\mathbf{W}^*$ as

$$\mathbf{W}^* = \underset{\mathbf{W} \in \mathcal{W}}{\arg\min} \left\| \mathbf{X}_1 - \mathbf{X}_0 \mathbf{W} \right\|. \tag{7}$$

Abadie et al. [2010] commented that we can select $\mathbf{W}^*$ so that (5) holds approximately and that weighted-adjustment estimation techniques of this form are not appropriate when the fit is poor. Note that $\mathbf{W}^*$ is not random since the covariates are assumed to be fixed. Since $\mathcal{W}$ is a closed and bounded subset of $\mathbb{R}^n$, $\mathcal{W}$ is compact. Because the objective function is continuous in $\mathbf{W}$, $\mathbf{W}^*$ will always exist. Our weighted-adjustment estimator for the shock effect $\alpha_1$ is

$$\hat{\alpha}_{\text{wadj}} = \sum_{i=2}^{n+1} w_i^* \hat{\alpha}_i \quad \text{for} \quad \mathbf{W}^* = \begin{pmatrix} w_2^* & \cdots & w_{n+1}^* \end{pmatrix}.$$

Estimation properties of $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{IVW}}$, and $\hat{\alpha}_{\text{wadj}}$ are discussed in the remaining sections.

**Remark 1.** In Section 2.1 we specify that $\mathbf{x}_{i,t}, \theta, \beta \in \mathbb{R}^p$. However, it is not necessary that the all $p$ covariates are important for every time series under study. The regression coefficients $\theta$ and $\beta$ are nuisance parameters that are not of primary importance. It will be understood that structural 0s in $\mathbf{x}_{i,t}$ correspond to variables that are unimportant.

# 3 Forecast risk and properties of shock-effects estimators

In this section, we discuss the properties that are related to forecast-risk reduction. In discussion of risk, it is useful to derive expressions for expectation and variance of the adjustment estimator $\hat{\alpha}_{\text{adj}}$ and weighted-adjustment estimator. The expression for the expectations are attached as follow[1].

(i) Under $\mathcal{M}_1$, $\mathrm{E}(\hat{\alpha}_{\text{adj}}) = \mathrm{E}(\hat{\alpha}_{\text{wadj}}) = \mu_\alpha$.

(ii) Under $\mathcal{M}_{21}$,

$$\mathrm{E}(\hat{\alpha}_{\text{adj}}) = \mu_\alpha + \frac{1}{2} \sum_{i=2}^{n+1} \delta' \mathbf{x}_{i,T_i^*} + \frac{1}{n} \sum_{i=2}^{n+2} \gamma' \mathbf{x}_{i,T_i^*-1} \quad \text{and} \quad \mathrm{E}(\hat{\alpha}_{\text{wadj}}) = \mu_\alpha + \delta' \mathbf{x}_{1,T_1^*} + \gamma' \mathbf{x}_{1,T_1^*-1}$$

(iii) Under $\mathcal{M}_{22}$,

$$\mathrm{E}(\hat{\alpha}_{\text{adj}}) = \mu_\alpha + \frac{1}{2} \sum_{i=2}^{n+1} \mu_\delta' \mathbf{x}_{i,T_i^*} + \frac{1}{n} \sum_{i=2}^{n+2} \mu_\gamma' \mathbf{x}_{i,T_i^*-1} \quad \text{and} \quad \mathrm{E}(\hat{\alpha}_{\text{wadj}}) = \mu_\alpha + \mu_\delta' \mathbf{x}_{1,T_1^*} + \mu_\gamma' \mathbf{x}_{1,T_1^*-1}$$

[1] The formal justification can be found in Section 5.1.1.

Note that $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ are not unbiased estimators for $\alpha_1$. Notice that under $\mathcal{M}_1$, $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{adj}}$ are unbiased estimators for $\text{E}(\alpha_1) = \mu_\alpha$ (see distributional details of $\alpha_1$ in Section 2.1). Nevertheless, $\hat{\alpha}_{\text{adj}}$ is a biased estimator for $\text{E}(\alpha_1)$ but $\hat{\alpha}_{\text{wadj}}$ is an unbiased estimator for $\text{E}(\alpha_1)$ under both $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$. Thus, we collect these results as the following proposition.

**Proposition 1.**

*(i) Under $\mathcal{M}_1$, $\hat{\alpha}_{\text{adj}}$ is an unbiased estimator of $\text{E}(\alpha_1)$. Under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$, $\hat{\alpha}_{\text{adj}}$ is a biased estimator of $\text{E}(\alpha_1)$ in general.*

*(ii) Suppose that $\mathbf{W}^*$ satisfies (5). Under $\mathcal{M}_1$, $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$, $\hat{\alpha}_{\text{wadj}}$ is an unbiased estimator of $\text{E}(\alpha_1)$.*

Unbiasedness properties for $\text{E}(\alpha_1)$ of $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$ allow for simple risk-reduction conditions and invoke a method of comparison, although our primary interest is in reducing forecast risk. These conditions will be discussed in Section 3.1 and Section 3.2. Next, we present the variance expressions for $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$ as below[2].

(i) Under $\mathcal{M}_1$ and $\mathcal{M}_{21}$,

$$\text{Var}(\hat{\alpha}_{\text{adj}}) = \frac{\sigma^2}{n^2} \sum_{i=2}^{n+1} \text{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \frac{\sigma_\alpha^2}{n^2}$$

$$\text{Var}(\hat{\alpha}_{\text{wadj}}) = \sigma^2 \sum_{i=2}^{n+1} (w_i^*)^2 \text{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \sigma_\alpha^2 \sum_{i=2}^{n+1} (w_i^*)^2$$

(ii) Under $\mathcal{M}_{22}$,

$$\text{Var}(\hat{\alpha}_{\text{adj}}) = \frac{\sigma^2}{n^2} \sum_{i=2}^{n+1} \text{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \frac{1}{n^2}\text{Var}(\alpha_i)$$

$$\text{Var}(\hat{\alpha}_{\text{wadj}}) = \sigma^2 \sum_{i=2}^{n+1} (w_i^*)^2 \text{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \sum_{i=2}^{n+1} (w_i^*)^2 \text{Var}(\alpha_i)$$

Note that the variances are not comparable in closed-form because of the term $\text{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\}$. This term exists because of the inclusion of the random lagged response in our auto regressive model formulation. Under $\mathcal{M}_{22}$, the expression for $\text{Var}(\alpha_i)$ is not of closed form because $\gamma_i$ and $\delta_i$ may be dependent when they are placed in a random-effects model. We investigate comparisons between the variability of these estimators in Section 3.2.

As Section 3.1 and 3.2 detailed the conditions for risk-reduction and comparisons, they usually involve fixed quantities related to variance and expectation. To make use of those properties in practice, estimation is required. Section **??** will introduce a general procedure of *parametric bootstrap* under the context of the problem to attain this purpose.

## 3.1 Conditions for risk-reduction for shock-effects estimators

In this section we will discuss the conditions for risk reduction for individual shock-effects estimators under $\mathcal{M}_1$, $\mathcal{M}_{21}$, and $\mathcal{M}_{22}$.

[2]The formal justification can be found in Section 5.1.2

### 3.1.1 Conditions under $\mathcal{M}_1$

Recall that Proposition 1 implies that the adjustment estimator $\hat{\alpha}_{\text{adj}}$ and weighted-adjustment estimator $\hat{\alpha}_{\text{wadj}}$ are unbiased for $\text{E}(\alpha_1)$ under $\mathcal{M}_1$. With this result, we will have the following propositions that specify the conditions that are necessary for risk reduction.

**Proposition 2.** *Under* $\mathcal{M}_1$,

*(i)* $R_{T_1^*+1,2} < R_{T_1^*+1,1}$ *when* $\text{Var}(\hat{\alpha}_{\text{adj}}) < \mu_\alpha^2$.

*(ii) if* $\mathbf{W}^*$ *satisfies (5),* $R_{T_1^*+1,2} < R_{T_1^*+1,1}$ *when* $\text{Var}(\hat{\alpha}_{\text{wadj}}) < \mu_\alpha^2$.

Proposition 2 tells that under $\mathcal{M}_1$ if the variance of the estimator is smaller than the squared mean of $\alpha_1$, those estimators will enjoy the risk reduction properties. Recalling from variance expression at the beginning of Section 3, Proposition 2 shows that the risk-reduction condition is

$$\text{Var}(\hat{\alpha}_{\text{adj}}) = \frac{\sigma^2}{n^2}\sum_{i=2}^{n+1}\text{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \frac{\sigma_\alpha^2}{n^2} < \mu_\alpha^2 \tag{8}$$

In terms of the adjustment estimator, $\hat{\alpha}_{\text{adj}}$, (8) implies two facts: (1) Forecast 2 is preferable to Forecast 1 asymptotically in $n$ whenever $\mu_\alpha \neq 0$; (2) In finite pool of time series, Forecast 2 is preferable to Forecast 1 when the $\mu_\alpha$ is large relative to its variability and overall regression variability.

For the weighted-adjustment estimator $\hat{\alpha}_{\text{wadj}}$, if $\mathbf{W}^*$ does not satisfy (5), its unbiased properties for $\text{E}(\alpha_1)$ should hold approximately when the fit in (7) is appropriate as commented in Section 2.3. From Proposition 2 and variance expression of $\hat{\alpha}_{\text{wadj}}$, the following is the risk-reduction condition for $\hat{\alpha}_{\text{wadj}}$.

$$\text{Var}(\hat{\alpha}_{\text{wadj}}) = \sigma^2\sum_{i=2}^{n+1}(w_i^*)^2\text{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \sigma_\alpha^2\sum_{i=2}^{n+1}(w_i^*)^2 < \mu_\alpha^2.$$

In this case, Forecast 2 is preferable to Forecast 1 when $\mu_\alpha$ is large relative to the *weighted sum of variances for shock effects for other time series* and overall regression variability. However, the above criteria are generally difficult to evaluate in practice due to the term $\hat{\alpha}_{\text{wadj}}$. Section **??** will provide a detailed treatment about how to deal with these technical inequalities in practice.

### 3.1.2 Conditions under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$

The $\alpha_i$s have different means under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$ unlike under $\mathcal{M}_1$. However, Proposition 1 implies that $\hat{\alpha}_{\text{wadj}}$ is an unbiased estimator of $\text{E}(\alpha_1)$. We now state conditions for risk reduction.

**Proposition 3.** *If* $\mathbf{W}^*$ *satisfies (5), under* $\mathcal{M}_{21}$ *and* $\mathcal{M}_{22}$, $R_{T_1^*+1,2} < R_{T_1^*+1,1}$ *when* $\text{Var}(\hat{\alpha}_{\text{wadj}}) < (\text{E}(\alpha_1))^2$.

Based on Proposition 3, we can obtain a similar inequality as in Section 3.1.1 as below

$$\text{Var}(\hat{\alpha}_{\text{wadj}}) = \sigma^2\sum_{i=2}^{n+1}(w_i^*)^2\text{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \sum_{i=2}^{n+1}(w_i^*)^2\text{Var}(\alpha_i) < (\text{E}(\alpha_1))^2,$$

where $\text{Var}(\alpha_i)$ may be replaced with $\sigma_\alpha^2$ in $\mathcal{M}_{21}$. The conclusions and intuitions will be identically the same as what we have in Section 3.1.1.

Proposition 1 shows that $\hat{\alpha}_{\text{adj}}$ is a biased estimator of $\text{E}(\alpha_1)$ under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$ generally. Hence, Proposition 2 no longer holds for $\hat{\alpha}_{\text{adj}}$ under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$. But, as an alternative, we can derive similar conditions as below. By Lemma 1 (see Section 5.1) and risk decomposition, we will achieve risk-reduction as long as

$$\begin{aligned}\text{E}(\alpha_1^2) = \text{Var}(\alpha_1) + (\text{E}(\alpha_1))^2 &> \text{E}(\hat{\alpha}_{\text{adj}} - \alpha_1)^2 \\ &= \text{Var}(\hat{\alpha}_{\text{adj}}) + (\text{E}(\hat{\alpha}_{\text{adj}}) - \alpha_1)^2 \\ &= \text{Var}(\hat{\alpha}_{\text{adj}}) + \text{Var}(\alpha_1) + (\text{E}(\hat{\alpha}_{\text{adj}}) - \text{E}(\alpha_1))^2\end{aligned}$$

Therefore, the above inequality will simply to

$$(\text{E}(\alpha_1))^2 > \text{Var}(\hat{\alpha}_{\text{adj}}) + (\text{E}(\hat{\alpha}_{\text{adj}}) - \text{E}(\alpha_1))^2.$$

Note that since $\hat{\alpha}_{\text{adj}}$ is biased for $\text{E}(\alpha_1)$, the bias term $(\text{E}(\hat{\alpha}_{\text{adj}}) - \text{E}(\alpha_1))^2$ will become complicated and simplification yields no insightful results.

As mentioned in Section 2.3, it is difficult to evaluate the expectation and variance of $\hat{\alpha}_{\text{IVW}}$. In other words, $\hat{\alpha}_{\text{IVW}}$ is generally biased for $\text{E}(\alpha_1)$. That is to say we can adapt the above proof to derive the risk-reduction conditions for $\hat{\alpha}_{\text{IVW}}$: under $\mathcal{M}_1$, $\mathcal{M}_{21}$, and $\mathcal{M}_{22}$, $R_{T_1^*+1,2} < R_{T_1^*+1,1}$ when $\text{Var}(\hat{\alpha}_{\text{IVW}}) + (\text{E}(\hat{\alpha}_{\text{IVW}}) - \text{E}(\alpha_1))^2 < (\text{E}(\alpha_1))^2$.

Topics of evaluation of these inequalities in practice can be found in Section **??**. As conditions of risk-reduction for individual estimators are obtained, a natural question arises — which one is better? It will be discussed in Section 3.2.

## 3.2 Comparisons among estimators

In comparing shock-effects estimators, we would assume that the risk-reduction conditions are satisfied as in Section 3.1.

Denote the risk-reduction quantity for the adjustment estimator as $\Delta_{\text{adj}}$, the one for inverse-weighted estimator as $\Delta_{\text{IVW}}$, and the one for weighted-adjustment estimator as $\Delta_{\text{wadj}}$. As long as the risk-reduction of one estimator is greater than those of others, we will vote it as the best estimator among our pool of estimators for consideration. For example, if we find that $\Delta_{\text{wadj}} > \Delta_{\text{adj}}$ and $\Delta_{\text{wadj}} > \Delta_{\text{IVW}}$, the weighted-adjustment estimator $\hat{\alpha}_{\text{wadj}}$ is the most favorable.

According to discussion in Section 3.1.2, we know that under $\mathcal{M}_1$, $\mathcal{M}_{21}$, and $\mathcal{M}_{22}$, the risk-reduction quantity for $\hat{\alpha}_{\text{IVW}}$ is

$$\Delta_{\text{IVW}} = (\text{E}(\alpha_1))^2 - \text{Var}(\hat{\alpha}_{\text{IVW}}) - (\text{E}(\hat{\alpha}_{\text{IVW}}) - \text{E}(\alpha_1))^2.$$

From discussions in Section 3.1, we know that the risk-reduction quantities for $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$ differ across models, we will discuss in different cases accordingly.

### 3.2.1 Under $\mathcal{M}_1$

From Proposition 2, we know that the risk-reduction quantities for $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$ are

$$\Delta_{\text{adj}} = \mu_\alpha^2 - \text{Var}(\hat{\alpha}_{\text{adj}}) \quad \text{and} \quad \Delta_{\text{wadj}} = \mu_\alpha^2 - \text{Var}(\hat{\alpha}_{\text{wadj}}).$$

Under the framework of $\mathcal{M}_1$, the risk-reduction quantity for $\hat{\alpha}_{\text{IVW}}$ is

$$\Delta_{\text{IVW}} = \mu_\alpha^2 - \text{Var}(\hat{\alpha}_{\text{IVW}}) - (\text{E}(\hat{\alpha}_{\text{IVW}}) - \mu_\alpha)^2.$$

In other words, when $\text{Var}(\hat{\alpha}_{\text{wadj}}) < \text{Var}(\hat{\alpha}_{\text{adj}})$ and $\hat{\alpha}_{\text{wadj}} < \text{Var}(\hat{\alpha}_{\text{IVW}}) + (\text{E}(\hat{\alpha}_{\text{IVW}}) - \mu_\alpha)^2$, we would prefer $\hat{\alpha}_{\text{wadj}}$ as the best estimator. Other conditions for voting the other estimators as the best one follow similarly.

### 3.2.2 Under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$

According to Proposition 3 and the discussion in Section 3.1.2, the risk-reduction quantities $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$ are

$$\Delta_{\text{adj}} = (\mathrm{E}(\alpha_1))^2 - \mathrm{Var}(\hat{\alpha}_{\text{adj}}) - (\mathrm{E}(\hat{\alpha}_{\text{adj}}) - \mathrm{E}(\alpha_1))^2 \quad \text{and} \quad \Delta_{\text{wadj}} = (\mathrm{E}(\alpha_1))^2 - \mathrm{Var}(\hat{\alpha}_{\text{wadj}}).$$

In this case, the risk-reduction quantity for $\hat{\alpha}_{\text{adj}}$ is similar to that of $\hat{\alpha}_{\text{IVW}}$ since they are both biased for $\mathrm{E}(\alpha_1)$. Thus,

$$\Delta_{\text{IVW}} = (\mathrm{E}(\alpha_1))^2 - \mathrm{Var}(\hat{\alpha}_{\text{IVW}}) - (\mathrm{E}(\hat{\alpha}_{\text{IVW}}) - \mathrm{E}(\alpha_1))^2$$

For the case of $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$ , we can derive the following inequality for $\hat{\alpha}_{\text{wadj}}$ to be favored over $\hat{\alpha}_{\text{adj}}$.

$$\mathrm{Var}(\hat{\alpha}_{\text{adj}}) - \mathrm{Var}(\hat{\alpha}_{\text{wadj}}) + \big(\mathrm{E}(\hat{\alpha}_{\text{adj}}) - \mathrm{E}(\alpha_1)\big)^2 > 0.$$

What insights can we gain from this inequality? We can analyze this inequality from two perspectives.

1. If it turns out to be fact that the variance of the weighted-adjustment estimator is greater than that of adjustment estimator, we should be aware that the compromise for variance because of using $\hat{\alpha}_{\text{wadj}}$ shouldn't exceed the squared bias, i.e., $\big(\mathrm{E}(\hat{\alpha}_{\text{adj}}) - \mathrm{E}(\alpha_1)\big)^2$.

2. If instead the variance of $\hat{\alpha}_{\text{wadj}}$ is smaller than that of $\hat{\alpha}_{\text{adj}}$, the above inequality should always hold because $\big(\mathrm{E}(\hat{\alpha}_{\text{adj}}) - \mathrm{E}(\alpha_1)\big)^2 > 0$ under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$.

These are some analytical results for comparison studies among estimators of $\alpha_1$. Next, we will detail a framework for estimation of risk-reduction quantities using non-parametric bootstrap such that the above inequalities can be analyzed numerically in practice.

## 3.3 Parametric bootstrap for risk-reduction evaluation problems

In this section, we present a parametric bootstrap procedure for our AR(1) model (see Section 2.1) in approximating the distribution of our shock-effect estimators. In the first step, we will introduce the pros and cons of this algorithm and refer the users to other alternatives. In the second step, we detail the procedure of boostrapping AR(1).

Efron [1979] introduces the bootstrapping method to approximate the distribution of a statistic using a *random sample*. However, this method will fail for time-series data since serial correlation or time-dependent features exist in typical time-series model.

In the usual setup of AR($p$) model, the unobserved errors are assumed to be *identically* and *independently distributed*. In other words, standard bootstrap methodology can be applied to resampling the residuals [Efron and Tibshirani, 1986]. As in the standard bootstrap, the asymptotic accuracy for OLS parameter estimation is guaranteed under the order of $O(T^{-1/2})$ almost surely, where $T$ is the length of the time series [Berkowitz and Kilian, 2000]. Bose [1988] showed that it can be further improved to $o(T^{-1/2})$ almost surely under some regularity conditions. Nevertheless, the pseudo time series generated by this procedure are not stationary.

Politis and Romano [1994] invented the stationary bootstrap method for *strictly stationary* and *weakly dependent* time series. However, the asymptotic accuracy of this procedure to OLS estimation is not known. Additionally, the asymptotic accuracy of this algorithm can be sensitive to the selection of $p$, the parameter of the geometric distribution; this issue is similar to that of the selection of block size in moving-block bootstrapping [Künsch, 1989, Liu et al., 1992]. More work related to bootstrapping time series can be referred to Berkowitz and Kilian [2000]. It is up to the user in selecting which procedure to choose but under *different* assumptions on the time-series.

The procedures of bootstrapping AR(1) can be outline as

1. From the observed fitted AR(1) model, we have a sample of residuals $\hat{\varepsilon}_{i,t}$'s for $i = 2, \ldots, n+1$ and $t = 1, \ldots, T_i$.
2. For the $b$th step of bootstrapping, we sampling with replacement from $\hat{\varepsilon}_{i,t}$'s to obtain the bootstrapped residuals $\hat{\varepsilon}^b_{i,t}$'s.
3. Pluging $\hat{\varepsilon}^*_{i,t}$ into the original model yields a sample of "bootstrapped" response $y^b_{i,t}$'s.
4. Compute the estimate of the shock-effects $\hat{\alpha}^b_i$ using OLS.
5. Compute the $\hat{\alpha}^b_{\text{adj}}$, $\hat{\alpha}^b_{\text{wadj}}$, and $\hat{\alpha}^b_{\text{IVW}}$.
6. Compute the sample mean, and sample variance of $\hat{\alpha}^b_{\text{adj}}$, $\hat{\alpha}^b_{\text{wadj}}$, and $\hat{\alpha}^b_{\text{IVW}}$.

Note that the estimates yielded by above procedure will provide an approximation for parameters involved in the risk-reduction conditions in Sections 3.1 and 3.2. In particular $\overline{\hat{\alpha}_{\text{wadj}}}$, the sample mean of $\hat{\alpha}^b_{\text{wadj}}$, will provide an approximation for $\mathrm{E}(\alpha_1)$ under $\mathcal{M}_1$, $\mathcal{M}_{21}$, and $\mathcal{M}_{22}$ since it is unbiased for $\mathrm{E}(\alpha_1)$ under those three configurations from Proposition 1. Therefore, we can judge whether to use a shock-effect estimator, and choose between shock-effect estimators by this method.

# 4 Simulation

# 5 Supplementary Materials

## 5.1 Proofs

### 5.1.1 Justification of Expectation of $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$

The building block for the following proof is the fact that least squares is conditionally unbiased conditioned on $\Theta$.

**Case I: under $\mathcal{M}_1$:** It follows that under $\mathcal{M}_1$ (see Section 2.1),

$$\mathrm{E}(\hat{\alpha}_{\text{adj}}) = \frac{1}{n}\sum_{i=2}^{n+1} \mathrm{E}(\mathrm{E}(\hat{\alpha}_i|\Theta)) = \mu_\alpha \quad \text{and} \quad \mathrm{E}(\hat{\alpha}_{\text{wadj}}) = \sum_{i=2}^{n+1} w_i^* \mathrm{E}(\mathrm{E}(\hat{\alpha}_i|\Theta)) = \sum_{i=2}^{n+1} w_i^* \mu_\alpha = \mu_\alpha.$$

where we used the fact that $\sum_{i=2}^{n+1} w_i = 1$.

**Case II: under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$:** Since $\mathrm{E}(\tilde{\varepsilon}_{i,T_i}) = 0$, $\mathrm{E}(\hat{\tilde{\alpha}}_i) = \mathrm{E}(\tilde{\alpha}_i) = \mathrm{E}(\alpha_i)$, it follows that

$$\begin{aligned}
\mathrm{E}(\hat{\alpha}_{\text{wadj}}) = \mathrm{E}\left\{\mathrm{E}\left(\sum_{i=2}^{n+1} w_i^* \hat{\alpha}_i|\Theta\right)\right\} &= \mathrm{E}\left(\sum_{i=2}^{n+1} w_i^* \alpha_i\right) \\
&= \mathrm{E}\Bigg\{\sum_{i=2}^{n+1} w_i^* \left[\mu_\alpha + \delta_i'\mathbf{x}_{i,T_i^*} + \gamma_i'\mathbf{x}_{i,T_i^*-1}\right]\Bigg\} \\
&= \mu_\alpha + \mathrm{E}\Bigg\{\sum_{i=2}^{n+1} w_i^* \left[\delta_i'\mathbf{x}_{i,T_i^*} + \gamma_i'\mathbf{x}_{i,T_i^*-1}\right]\Bigg\}. \qquad (\mathbf{W} \in \mathcal{W})
\end{aligned}$$

Similarly,

$$\mathrm{E}(\hat{\alpha}_{\mathrm{adj}}) = \mu_\alpha + \frac{1}{n}\sum_{i=2}^{n+1}\mathrm{E}(\delta_i'\mathbf{x}_{i,T_i^*} + \gamma_i'\mathbf{x}_{i,T_i^*-1}).$$

### 5.1.2 Justification of Variance of $\hat{\alpha}_{\mathrm{adj}}$ and $\hat{\alpha}_{\mathrm{wadj}}$

Notice that under the setting of OLS, the design matrix for $\mathcal{M}_2$ is the same as the one for $\mathcal{M}_1$. Therefore, it follows that

$$\begin{aligned}\mathrm{Var}(\hat{\alpha}_{\mathrm{wadj}}) &= \mathrm{E}(\mathrm{Var}(\hat{\alpha}_{\mathrm{wadj}}|\Theta)) + \mathrm{Var}(\mathrm{E}(\hat{\alpha}_{\mathrm{wadj}}|\Theta)) \\ &= \mathrm{E}\left\{\mathrm{Var}\left(\sum_{i=2}^{n+1} w_i^*\hat{\alpha}_i|\Theta\right)\right\} + \mathrm{Var}\left(\sum_{i=2}^{n+1} w_i^*\alpha_i\right)\end{aligned}$$

Under $\mathcal{M}_{21}$ where $\delta_i = \delta$ and $\gamma_i = \gamma$ are fixed unknown parameters, we will have

$$\begin{aligned}\mathrm{Var}(\hat{\alpha}_{\mathrm{wadj}}) &= \mathrm{E}\left\{\sum_{i=2}^{n+1}(w_i^*)^2(\sigma^2(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22})\right\} + \sigma_\alpha^2\sum_{i=2}^{n+1}(w_i^*)^2 \\ &= \sigma^2\sum_{i=2}^{n+1}(w_i^*)^2\mathrm{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \sigma_\alpha^2\sum_{i=2}^{n+1}(w_i^*)^2. \end{aligned} \tag{9}$$

Similarly, under $\mathcal{M}_{22}$ where we assume $\delta_i \perp\!\!\!\perp \gamma_i \perp\!\!\!\perp \varepsilon_{i,t}$, we have

$$\mathrm{Var}(\hat{\alpha}_{\mathrm{wadj}}) = \sigma^2\sum_{i=2}^{n+1}(w_i^*)^2\mathrm{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \sum_{i=2}^{n+1}(w_i^*)^2\mathrm{Var}(\alpha_i)$$

For the adjustment estimator, we simply replace $\mathbf{W}^*$ with $1/n\mathbf{1}_n$. Thus, under $\mathcal{M}_{21}$ we have

$$\mathrm{Var}(\hat{\alpha}_{\mathrm{adj}}) = \frac{\sigma^2}{n^2}\sum_{i=2}^{n+1}\mathrm{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \frac{\sigma_\alpha^2}{n^2}$$

Under $\mathcal{M}_{22}$, we shall have

$$\mathrm{Var}(\hat{\alpha}_{\mathrm{adj}}) = \frac{\sigma^2}{n^2}\sum_{i=2}^{n+1}\mathrm{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \frac{1}{n^2}\mathrm{Var}(\alpha_i).$$

Notice that $\mathcal{M}_1$ differs from $\mathcal{M}_{21}$ only by its mean parameterization of $\alpha$ (see Section 2.1). In other words, the variances of $\hat{\alpha}_{\mathrm{adj}}$ and $\hat{\alpha}_{\mathrm{wadj}}$ under $\mathcal{M}_1$ are the same for those under $\mathcal{M}_{21}$.

## 5.2 Proofs for lemmas and propositions

**Proof of Proposition 1** The proof for unbiasedness follows immediately from discussions related to expectation in Section 3. For the biasedness of $\hat{\alpha}_{\mathrm{adj}}$ under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$, we write the bias term for $\hat{\alpha}_{\mathrm{adj}}$ as below.

$$\mathrm{Bias}(\hat{\alpha}_{\mathrm{adj}}) = \begin{cases}\frac{1}{n}\sum_{i=2}^{n+1}\delta'(\mathbf{x}_{i,T_i^*} - n\mathbf{x}_{1,T_1^*}) + \frac{1}{n}\sum_{i=2}^{n+1}\gamma'(\mathbf{x}_{i,T_i^*-1} - n\mathbf{x}_{1,T_1^*-1}) & \text{for } \mathcal{M}_{21} \\ \frac{1}{n}\sum_{i=2}^{n+1}\mu_\delta'(\mathbf{x}_{i,T_i^*} - n\mathbf{x}_{1,T_1^*}) + \frac{1}{n}\sum_{i=2}^{n+1}\mu_\gamma'(\mathbf{x}_{i,T_i^*-1} - n\mathbf{x}_{1,T_1^*-1}) & \text{for } \mathcal{M}_{22}\end{cases}.$$

But it may be unbiased in some special circumstances when the above bias turns out to be 0. $\square$

**Lemma 1.** *The forecast risk difference is* $R_{T_1^*+1,1} - R_{T_1^*+1,2} = \mathrm{E}(\alpha_1^2) - \mathrm{E}(\hat{\alpha} - \alpha_1)^2$ *for all estimators of* $\alpha_1$ *that are independent of* $\Theta_1$ *(see Section 2.1).*

**Proof of Lemma 1** Define

$$C(\Theta_1) = \hat{\eta}_1 + \hat{\phi}_1 y_{1,T_1^*} + \hat{\theta}_1' \mathbf{x}_{1,T_1^*+1} + \hat{\beta}_1' \mathbf{x}_{1,T_1^*} - (\eta_1 + \phi_1 y_{1,T_1^*} + \theta_1' \mathbf{x}_{1,T_1^*+1} + \beta_1' \mathbf{x}_{1,T_1^*}),$$

where $\Theta_1$ is as defined in (3). Notice that

$$R_{T_1^*+1,1} = \mathrm{E}\big\{\big(C(\Theta_1) - \alpha_1\big)^2\big\} \qquad \text{and} \qquad R_{T_1^*+1,2} = \mathrm{E}\big\{\big(C(\Theta_1) + \hat{\alpha} - \alpha_1\big)^2\big\}.$$

It follows that

$$R_{T_1^*+1,1} - R_{T_1^*+1,2} = \mathrm{E}(\alpha_1^2) - 2\mathrm{E}(C(\Theta_1)\hat{\alpha}) - \mathrm{E}(\hat{\alpha} - \alpha_1)^2.$$

Assuming $\mathbf{S} = (\mathbf{1}_n, \mathbf{y}_{1,t-1}, \mathbf{x}_1, \mathbf{x}_{1,t-1})$ has full rank, under OLS setting, $\hat{\eta}_1$, $\hat{\phi}_1$, $\hat{\theta}_1$, and $\hat{\beta}_1$ are unbiased estimators of $\eta_1$, $\phi_1$, $\theta_1$, and $\beta_1$, respectively under conditioning of $\Theta_1$. Since we assume $\hat{\alpha}$ is independent of $\Theta_1$, through the method of iterated expectation,

$$\mathrm{E}(C(\Theta_1)\hat{\alpha}) = \mathrm{E}\big\{\hat{\alpha} \cdot \mathrm{E}(C(\Theta_1) \mid \Theta_1)\big\} = 0.$$

It follows that

$$R_{T_1^*+1,1} - R_{T_1^*+1,2} = \mathrm{E}(\alpha_1^2) - \mathrm{E}(\hat{\alpha} - \alpha_1)^2,$$

which finishes the proof. □

**Proof of Proposition 2** The proofs are arranged into two separate parts as below.

**Proof for statement (i):** Under $\mathcal{M}_1$, $\hat{\alpha}_{\text{adj}}$ is an unbiased estimator of $\mathrm{E}(\alpha_1)$ because

$$\begin{aligned} E\left(\frac{1}{n}\sum_{i=2}^{n+1}\hat{\alpha}_i\right) = \frac{1}{n}\sum_{i=2}^{n+1}\mathrm{E}(\hat{\alpha}_i) &= \frac{1}{n}\sum_{i=2}^{n+1}\mathrm{E}(\mathrm{E}(\hat{\alpha}_i \mid \Theta)) \\ &= \frac{1}{n}\sum_{i=2}^{n+1}\mathrm{E}(\alpha_i) = \mu_\alpha = \mathrm{E}(\alpha_1), \end{aligned}$$

where we used the fact that OLS estimator is unbiased when the design matrix $\mathbf{U}_i$ is of full rank for all $i = 2, \ldots, n+1$. Because $\alpha_1 \perp\!\!\!\perp \varepsilon_{i,t}$, $\mathrm{E}(\hat{\alpha}_{\text{adj}}\alpha_1) = \mathrm{E}(\hat{\alpha}_{\text{adj}})\mathrm{E}(\alpha_1) = (\mathrm{E}(\hat{\alpha}_{\text{adj}}))^2$. By Lemma 1,

$$\begin{aligned} R_{T_1^*+1,1} - R_{T_1^*+1,2} &= \mathrm{E}(\alpha_1^2) - \mathrm{E}(\hat{\alpha}_{\text{adj}} - \alpha_1)^2 \\ &= \mathrm{E}(\alpha_1^2) - \mathrm{E}(\alpha_1^2) - \mathrm{E}(\hat{\alpha}_{\text{adj}}^2) + 2\mathrm{E}(\hat{\alpha}_{\text{adj}}\alpha_1) \\ &= \mu_\alpha^2 - \mathrm{Var}(\hat{\alpha}_{\text{adj}}) \end{aligned}$$

Therefore, as long as we have $\mathrm{Var}(\hat{\alpha}_{\text{adj}}) < \mu_\alpha^2$, we will achieve the risk reduction.

**Proof for statement (ii):** By Proposition 1, the property that $\hat{\alpha}_{\text{wadj}}$ is an unbiased estimator of $\mu_\alpha$ holds for $\mathcal{M}_1$. Then, similar proof of statement (i) will proves the Proposition. □

**Proof of Proposition 3** By Proposition 1, the property that $\hat{\alpha}_{\text{wadj}}$ is an unbiased estimator of $\mathrm{E}(\alpha_1)$ holds for $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$. Then, similar proof of Proposition **??** will proves the Proposition. □

## 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.

John M Bates and Clive WJ Granger. The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969.

Jeremy Berkowitz and Lutz Kilian. Recent developments in bootstrapping time series. *Econometric Reviews*, 19(1):1–48, 2000.

Richard Blundell and Stephen Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143, 1998.

Arup Bose. Edgeworth correction by bootstrap in autoregressions. *The Annals of Statistics*, pages 1709–1722, 1988.

B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7:1–26, 1979.

Bradley Efron and Robert Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, pages 54–75, 1986.

Clive William John Granger and Paul Newbold. *Forecasting economic time series*. Academic Press, 2014.

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

Se Yoon Lee, Bowen Lei, and Bani K. Mallick. Estimation of covid-19 spread curves integrating global data and borrowing information, 2020.

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

Yair Mundlak. On the pooling of time series and cross section data. *Econometrica: Journal of the Econometric Society*, pages 69–85, 1978.

Mogens Graf Plessen. Integrated time series summarization and prediction algorithm and its application to covid-19 data mining, 2020.

Dimitris N Politis and Joseph P Romano. The stationary bootstrap. *Journal of the American Statistical association*, 89(428):1303–1313, 1994.

Allan Timmermann. Forecast combinations. *Handbook of Economic Forecasting*, 1:135–196, 2006.

Arnold Zellner, Chansik Hong, and Chung-ki Min. Forecasting turning points in international output growth rates using bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques. *Journal of Econometrics*, 49(1-2):275–304, 1991.