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. Data analysis

Updates in this version

1. Figure 1 is updated.

2. Some introductions for forecast combination are added in Paragraph 2. Literature review for time-series pooling in cross-sectional (panel) data is rewritten and more to be added.

3. The procedure for parametric bootstrap is rewritten.

4. Technical details for SCM method are updated.

5. SCM code is written using `Rsolnp` package.

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 22, 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

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. However, all is not lost in this setting, one can integrate information from disparate time series that have previously undergone similar structural shocks to estimate the shock effect of the time series under study. One can then combine these past similar shock effects and add them to the present forecast to reduce the overall forecast error.

Improving forecasts through forecast combination has a rich history [Bates and Granger, 1969, Mundlak, 1978, Timmermann, 2006, Granger and Newbold, 2014]. The classical setting for the forecast combination problem is when there are competing forecasts for a single time series. **The following list of methods needs to fall in the forecast combination literature, not the time series pooling literature**. In this setting there are a plethora of methods for combining forecasts, e.g., (1) model averaging [Newbold and Harvey, 2002, Timmermann, 2006, Hansen, 2008], (2) model selection [Lee and Phillips, 2015, Greenaway-McGrevy, 2020], (3) time-series pooling [Mundlak, 1978, Zellner et al., 1991, Lee et al., 2020, Plessen, 2020]. Model averaging typically selects weights for models based on minimizing various loss functions whereas model selection chooses the model through minimization of those loss functions, see for example, Fosten and Greenaway-McGrevy [2019] **I do not think that this reference belongs here**. Classical forecast combination may fail when forecasting in the presence of structural shocks.

[*]jileil2@ilinois.edu

[†]ziyuliu3@illinois.edu

[‡]dje13@illinois.edu

In our post-shock setting we combine estimated quantities from different time series with the aim of lowering forecast error for a single time series under study. Our techniques are similar to those in time-series pooling and data integration. The literature of time-series pooling is mainly related to pooling cross-sectional panel data [Mundlak, 1978, Zellner et al., 1991, Fosten and Greenaway-McGrevy, 2019]. The issues about whether to assume homogeneity or heterogeneity of slope coefficients across individual units are confounded. Baltagi [2008] showed that homogeneity approach often outperform heterogeneity one in mean squared forecast error; while heterogeneity approach is more general to accommodate differences among units. **How does this relevant to what we are doing?**

Data integration for forecasting is a broad area of research including ideas from many areas. Lee et al. [2020] constructed a Bayesian hierarchical model embracing data integration to improve predictive precision of COVID-19 infection trajectories for different countries. A similar setup may be beneficial for post-shock prediction but may be too dependent upon model specification for the shock distribution. Plessen [2020] employed a data-mining approach to combine COVID-19 data from different countries as input to predict global net daily infections and deaths of COVID-19 using clustering. However, there is a tremendous amount of volatility in this form of COVID-19 data, and the fit of this prediction method may be improved with modeling structure or preprocessing of the donor pool. **Mention synthetic intervention paper here, stress the importance of carefully constructing the donor pool.** *More to be added ...*

Need more of a transition to what we do here and why it is good. 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 similar to that in Blundell and Bond [1998] with a general random effects structure. The main 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].

Need to discuss our example and our simulation results.

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 dynamic panel data model with autoregressive structure similar 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.

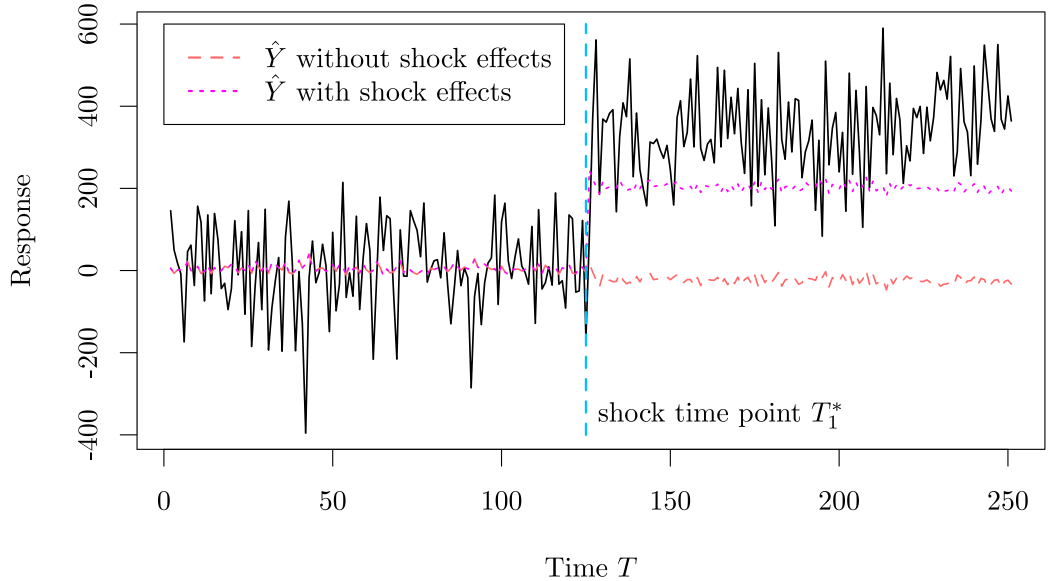

Figure 1. A comparison between forecast without considering shock effects and the one uses simple averaging given $n = 40$ disparate time series and that the shock time is at $T_1^* = 125$. The dots represent least square estimate $\hat{\alpha}_i$ from disparate time series. Two model differs by an adjustment $\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_i$. **The picture needs to be made more clear. Remove $\hat{\alpha}_i$ and state what those points are in the caption. Label the true (unknown to us) shock effect α_1.**

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:

$$\eta_i \overset{iid}{\sim} \eta, \text{ where } \mathrm{E}(\eta) = 0, \mathrm{Var}(\eta) = \sigma_\eta^2, \qquad i = 1, \ldots, n+1,$$
$$\phi_i \overset{iid}{\sim} \phi, \text{ where } |\phi| < 1, \qquad i = 1, \ldots, n+1,$$

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

Notice that $\mathcal{M}_1$ assumes that α_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:

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

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 δ_i and γ_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 γ and δ are fixed unknown parameters. We let $\mathcal{M}_{22}$ denote model $\mathcal{M}_2$ with the following random effects structure for γ and δ:

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

Note that δ_i and γ_i may be dependent. We further define the parameter sets

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

where Θ and Θ_1 can adapt to $\mathcal{M}_1$ by dropping δ_i and γ_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 η_1, ϕ_1, θ_1, and β_1 respectively, and $\hat{\alpha}$ is some form of estimator for the shock effect of time series of interest, i.e., α_1. The first forecast ignores the presence of α_1 while the second forecast incorporates an estimate of α_1 that is obtained from the other independent 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 α_1 nor does it converge to α_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 α_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 α_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 α_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 α_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 α_is. We can use $\hat{\alpha}_{\text{adj}}$ as an estimator for the unknown α_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 α_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

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

$\hat{\alpha}_i$ is the OLS estimator of α_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 σ 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 $\mathrm{E}(\hat{\alpha}_{\text{IVW}})$ and $\mathrm{Var}(\hat{\alpha}_{\text{IVW}})$, 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

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{x}_{1,T_1^*-1} \\ \mathbf{x}_{1,T_1^*} \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{X}}_1(\mathbf{W}) = w_2 \begin{pmatrix} \mathbf{x}_{2,T_2^*-1} \\ \mathbf{x}_{2,T_2^*} \end{pmatrix} + \cdots + w_{n+1} \begin{pmatrix} \mathbf{x}_{n+1,T_{n+1}^*} \\ \mathbf{x}_{n+1,T_{n+1}^*} \end{pmatrix}.$$

where $\mathbf{X}_1$ and $\hat{\mathbf{X}}_1(\mathbf{W})$ are $2 \times p$. 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

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

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

$$\left\{ \begin{pmatrix} \mathbf{x}_{2,T_2^*-1} \\ \mathbf{x}_{2,T_2^*} \end{pmatrix}, \ldots, \begin{pmatrix} \mathbf{x}_{n+1,T_{n+1}^*} \\ \mathbf{x}_{n+1,T_{n+1}^*} \end{pmatrix} \right\}.$$

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

$$\mathbf{W}^* = \underset{\mathbf{W} \in \mathcal{W}}{\arg\min} \left\| \mathrm{vec}\big(\mathbf{X}_1 - \hat{\mathbf{X}}_1(\mathbf{W})\big) \right\|_{2p}. \tag{6}$$

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 α_1 is

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

Estimation properties of $\hat{\alpha}_{\mathrm{adj}}$, $\hat{\alpha}_{\mathrm{IVW}}$, and $\hat{\alpha}_{\mathrm{wadj}}$ 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 θ and β 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 $\hat{\alpha}_{\mathrm{adj}}$ and weighted-adjustment estimator. The expression for the expectations are as follow,

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

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

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

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

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

Formal justification for these results can be found in Appendix. Note that $\hat{\alpha}_{\mathrm{adj}}$, $\hat{\alpha}_{\mathrm{wadj}}$, and $\hat{\alpha}_{\mathrm{IVW}}$ are not unbiased estimators for α_1. Notice that under $\mathcal{M}_1$, $\hat{\alpha}_{\mathrm{adj}}$ and $\hat{\alpha}_{\mathrm{adj}}$ are unbiased estimators for $\mathrm{E}(\alpha_1) = \mu_\alpha$ (see distributional details of α_1 in Section 2.1). Nevertheless, $\hat{\alpha}_{\mathrm{adj}}$ is a biased estimator for $\mathrm{E}(\alpha_1)$ but $\hat{\alpha}_{\mathrm{wadj}}$ is an unbiased estimator for $\mathrm{E}(\alpha_1)$ under both $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$. Thus, we collect these results as the following proposition.

Proposition 1.

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

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

Unbiasedness properties for $\mathrm{E}(\alpha_1)$ of $\hat{\alpha}_{\mathrm{adj}}$ and $\hat{\alpha}_{\mathrm{wadj}}$ 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 $\hat{\alpha}_{\mathrm{adj}}$ and $\hat{\alpha}_{\mathrm{wadj}}$ as below

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

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

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

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

Formal justification for these results can be found in Appendix. Note that the variances are not comparable in closed-form because of the term $\mathrm{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\}$. 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 $\mathrm{Var}(\alpha_i)$ is not of closed form because γ_i and δ_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 3.3 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 $\mathcal{M}_{22}$.

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 $\mathrm{E}(\alpha_1)$ 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* $\mathrm{Var}(\hat{\alpha}_{\text{adj}}) < \mu_\alpha^2$.

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

Proposition 2 tells that under $\mathcal{M}_1$ if the variance of the estimator is smaller than the squared mean of α_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

$$\mathrm{Var}(\hat{\alpha}_{\text{adj}}) = \frac{\sigma^2}{n^2}\sum_{i=2}^{n+1}\mathrm{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)_{22}^{-1}\big\} + \frac{\sigma_\alpha^2}{n^2} < \mu_\alpha^2 \tag{7}$$

In terms of the adjustment estimator, $\hat{\alpha}_{\text{adj}}$, (7) 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 μ_α 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 $\mathrm{E}(\alpha_1)$ should hold approximately when the fit in (6) 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 $\hat{\alpha}_{\text{wadj}}$.

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

In this case, Forecast 2 is preferable to Forecast 1 when μ_α 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 3.3 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 α_is have different means under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$ unlike under $\mathcal{M}_1$. However, Proposition 1 implies that $\hat{\alpha}_{\mathrm{wadj}}$ is an unbiased estimator of $\mathrm{E}(\alpha_1)$. 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* $\mathrm{Var}(\hat{\alpha}_{\mathrm{wadj}}) < (\mathrm{E}(\alpha_1))^2$.

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

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

where $\mathrm{Var}(\alpha_i)$ may be replaced with σ_α^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 $\hat{\alpha}_{\mathrm{adj}}$ is a biased estimator of $\mathrm{E}(\alpha_1)$ under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$ generally. Hence, Proposition 2 no longer holds for $\hat{\alpha}_{\mathrm{adj}}$ 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

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

Therefore, the above inequality will simply to

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

Note that since $\hat{\alpha}_{\mathrm{adj}}$ is biased for $\mathrm{E}(\alpha_1)$, the bias term $(\mathrm{E}(\hat{\alpha}_{\mathrm{adj}}) - \mathrm{E}(\alpha_1))^2$ 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 $\hat{\alpha}_{\mathrm{IVW}}$. In other words, $\hat{\alpha}_{\mathrm{IVW}}$ is generally biased for $\mathrm{E}(\alpha_1)$. That is to say we can adapt the above proof to derive the risk-reduction conditions for $\hat{\alpha}_{\mathrm{IVW}}$: under $\mathcal{M}_1$, $\mathcal{M}_{21}$, and $\mathcal{M}_{22}$, $R_{T_1^*+1,2} < R_{T_1^*+1,1}$ when $\mathrm{Var}(\hat{\alpha}_{\mathrm{IVW}}) + (\mathrm{E}(\hat{\alpha}_{\mathrm{IVW}}) - \mathrm{E}(\alpha_1))^2 < (\mathrm{E}(\alpha_1))^2$.

Topics of evaluation of these inequalities in practice can be found in Section 3.3. We will discuss comparisons of adjustment estimators in the the next Section.

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 Δ_{adj}, the one for inverse-weighted estimator as Δ_{IVW}, and the one for weighted-adjustment estimator as Δ_{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 $\Delta_{\mathrm{wadj}} > \Delta_{\mathrm{adj}}$ and $\Delta_{\mathrm{wadj}} > \Delta_{\mathrm{IVW}}$, the weighted-adjustment estimator $\hat{\alpha}_{\mathrm{wadj}}$ 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 $\hat{\alpha}_{\mathrm{IVW}}$ is

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

From discussions in Section 3.1, we know that the risk-reduction quantities for $\hat{\alpha}_{\mathrm{adj}}$ and $\hat{\alpha}_{\mathrm{wadj}}$ 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

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

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 $\text{E}(\alpha_1)$. Thus,

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

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 $\hat{\alpha}_{\text{adj}}$.

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

We 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., $\big(\text{E}(\hat{\alpha}_{\text{adj}}) - \text{E}(\alpha_1)\big)^2$.

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 $\big(\text{E}(\hat{\alpha}_{\text{adj}}) - \text{E}(\alpha_1)\big)^2 > 0$ under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$.

These are some analytical results for comparison studies among estimators of α_1. Next, we will detail a framework for estimation of risk-reduction quantities using a parametric bootstrap routine. Therefore, the above inequalities can be analyzed numerically in practice.

3.3 Parametric bootstrap for risk-reduction evaluation problems

I am not sure if all of this is necessary. This is a nice summary of the literature but I think that we only need the results that motivate our bootstrap procedure.

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 AR(p) model, the unobserved errors are assumed to be identically and independently distributed. In this setting, standard bootstrap methodology can be applied to resampling the residuals [Efron and Tibshirani, 1986]. 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] motivated a 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 (**what does this mean, what is the geometric distribution in this context?**). 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.

We do not need to rewrite bootstrap algorithms that exist in the literature. We only need to write our algorithm. Emphasis needs to be placed on T_i^*being known, the pre and post T_i^* sampling needs to be made explicit.

Algorithm 1: Parametric bootstrap for approximation for mean and variance of shock-effect estimators of α_1.

Input: B – the number of parametric bootstraps
$S_i = \{\hat{\varepsilon}_{i,t} \colon t = 1, \ldots, T_i\}$ – the collection of residuals for $t = 1, \ldots, T_i$
$\mathcal{P}_i$ – the fitted AR(1) model for $i = 2, \ldots, n+1$

Result: The sample mean, and sample variance of bootstrapped adjustment estimator, inverse-variance weighted estimator, and weighted-adjustment estimator.

for $b = 1 : B$ **do**
  for $i = 2, \ldots, n+1$ **do**
    Sample with replacement from S_i to obtain $S_i^{(b)} := \{\hat{\varepsilon}_{i,t}^{(b)} \colon t = 1, \ldots, T_i\}$
    Plug $S_i^{(b)}$ into $\mathcal{P}_i$ to generate the bootstrapped response $\mathbf{Y}_i^{(b)}$
    Compute $\hat{\alpha}_i^{(b)}$ based on OLS estimation of AR(1) with $\mathbf{Y}_i^{(b)}$ and covariates from $\mathcal{P}_i$
  end
  Compute $\hat{\alpha}_{\text{adj}}^{(b)}$, $\hat{\alpha}_{\text{wadj}}^{(b)}$, and $\hat{\alpha}_{\text{IVW}}^{(b)}$
end
Compute the sample mean, and sample variance of $\hat{\alpha}_{\text{adj}}^{(b)}$, $\hat{\alpha}_{\text{wadj}}^{(b)}$, and $\hat{\alpha}_{\text{IVW}}^{(b)}$.

The procedures of bootstrapping AR(1) can be outline as in Algorithm 1. We stress that the $\varepsilon_{i,t}$ and $\varepsilon_{i,t'}$ for $t \in \{1, \ldots, T_i^*\}$ and $t' \in \{T_i^* + 1, \ldots, T_i\}$ are i.i.d. since $y_{i,t}$ differs from $y_{i,t'}$ only in the shock-effect α_i conditioned on Θ (see Section 2.1), and it is *not* absorbed into $\varepsilon_{i,t}$ and $\varepsilon_{i,t'}$. Therefore, this parametric bootstrap framework will work for us.

The estimates yielded by Algorithm 1 provides 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 $\hat{\alpha}_{\text{wadj}}^{b}$, 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.

We should include a brief simulation which shows that our bootstrap procedure works.

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 Θ.

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,

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

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

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

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

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

$$\begin{aligned}
\mathrm{Var}(\hat{\alpha}_{\text{wadj}}) &= \mathrm{E}\left\{\sum_{i=2}^{n+1}(w_i^*)^2(\sigma^2(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22})\right\} + \sigma_\alpha^2\sum_{i=2}^{n+1}(w_i^*)^2 \\
&= \sigma^2\sum_{i=2}^{n+1}(w_i^*)^2\mathrm{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \sigma_\alpha^2\sum_{i=2}^{n+1}(w_i^*)^2. \qquad (8)
\end{aligned}$$

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 α (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 α_1 that are independent of Θ_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 Θ_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 η_1, ϕ_1, θ_1, and β_1, respectively under conditioning of Θ_1. Since we assume $\hat{\alpha}$ is independent of Θ_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 $\mathcal{M}_1$, $\hat{\alpha}_{\mathrm{adj}}$ 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 $\alpha_1 \perp\!\!\!\perp \varepsilon_{i,t}$, $\mathrm{E}(\hat{\alpha}_{\mathrm{adj}}\alpha_1) = \mathrm{E}(\hat{\alpha}_{\mathrm{adj}})\mathrm{E}(\alpha_1) = (\mathrm{E}(\hat{\alpha}_{\mathrm{adj}}))^2$. By Lemma 1,

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

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

Proof for statement (ii): By Proposition 1, the property that $\hat{\alpha}_{\mathrm{wadj}}$ is an unbiased estimator of μ_α holds for $\mathcal{M}_1$. The remainder of the proof follows a similar argument to the proof of statement (i). □

Proof of Proposition 3 By Proposition 1, the property that $\hat{\alpha}_{\mathrm{wadj}}$ is an unbiased estimator of $\mathrm{E}(\alpha_1)$ holds for $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$. The remainder of the proof follows a similar argument to the proof of Proposition **??** (**Fix this problematic reference**). □

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.

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

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.

Jack Fosten and Ryan Greenaway-McGrevy. Panel data nowcasting. *Available at SSRN 3435691*, 2019.

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

Ryan Greenaway-McGrevy. Multistep forecast selection for panel data. *Econometric Reviews*, 39 (4):373–406, 2020.

Bruce E Hansen. Least-squares forecast averaging. *Journal of Econometrics*, 146(2):342–350, 2008.

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.

Yoonseok Lee and Peter CB Phillips. Model selection in the presence of incidental parameters. *Journal of Econometrics*, 188(2):474–489, 2015.

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.

Paul Newbold and David I Harvey. Forecast combination and encompassing. *A companion to economic forecasting*, 1:620, 2002.

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.