Note

Updates in this version

1. Revisions to abstract, introduction, simulation, data analysis, and discussions.
2. Addition of Section 3.4
3. A major addition of content to simulation.

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

July 6, 2020

Abstract

We developed a forecasting methodology for providing credible forecasts for time series data that has undergone a shock by borrowing knowledge from disparate time series that have undergone similar shocks for which post-shock outcome is recorded. Three shock-effects estimators were constructed for minimizing average forecast risk. We proposed risk-reduction propositions providing conditions when our methodology works and estimators for risk-reduction quantities. Bootstrap procedures are provided to estimate the variability of our shock effect estimators; and these procedures can be used to assess the potential success of post-shock prediction before the post-shock response is observed. The risk-reduction propositions and risk-reduction quantities are powerful tools for users because we can empirically aid a prospective evaluation about whether the three aggregation techniques will work well and which one is the best. Leave-one-out cross validation with k random draws is proposed to estimate consistency of risk-reduction propositions and best consistency of voting the best shock-estimators, prospectively informing users the probabilities that this prospective evaluation is consistent with the reality. Several simulated data examples, and a real data example of forecasting Conoco Phillips stock price are provided for verification and illustration.

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. Then, we discuss some frequently used methodologies that may be applied in this situation.

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. In this setting there are a plethora of methods for combining forecasts, e.g., model averaging [Newbold and Harvey, 2002, Hendry and Clements, 2004, Koop and Potter, 2004, Timmermann, 2006, Eklund and Karlsson, 2007, Hansen, 2008], model selection [Swanson and White, 1997, Swanson and Zeng, 2001, Lee and Phillips, 2015], and other various methods [Pesaran and Pick, 2011, Li and Chen, 2014]. Forecast encompassing methods are

[*]jileil2@ilinois.edu
[†]ziyuliu3@illinois.edu
[‡]dje13@illinois.edu

developed to test whether competing forecasts are suitable for combinations [Newbold and Harvey, 2002, Fang, 2003].

As alternatives to forecast combination, data integration and time series pooling for forecasting is a broad domain of research including ideas from many areas. Time-series pooling methods are frequently used for panel data [Zellner et al., 1991, Hoogstrate et al., 2000, Baltagi, 2008, Fosten and Greenaway-McGrevy, 2019, Liu et al., 2020]. Recent researches in data integration have investigated versatile methods to combine knowledge from multiple datasets for better prediction, whose ideas are similar to ours. For instance, 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. Agarwal et al. [2020] proposed a model-free synthetic intervention method to predict unobserved potential outcomes after different interventions given a donor pool of observed outcomes with given interventions. They also provide useful guidelines for how to estimate the effects of potential interventions by giving recommendations for choosing the metric of interest, the intervention of interest, time horizons, and the donor pool.

Up to the scale of the shock, it is very unlikely that the above mentioned methods will work ideally since they are trained on the time series data that do not experience such a shock. To combat this problem, we develop and compare aggregation techniques in this post-shock setting based on the idea of data integration. 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 provide a scalar adjustment, based on estimated shock effects from the disparate time series, to the original forecast at the known shock time point.

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]. We provide risk-reduction propositions that detail the conditions when the adjusted forecasts will work better than the original one, and estimate risk-reduction quantities to find the best technique out of the three. The involved parameters in the risk-reduction propositions and risk-reduction quantities are estimated by parametric bootstrap. These propositions and risk-reduction quantities are powerful tools for users in the sense that we can empirically aid a prospective evaluation about whether the three aggregation techniques will work well and which one is the best, using estimates from bootstrap procedures. Furthermore, to inform the credibility of this prospective evaluation, we propose a leave-one-out cross validation with k random draws to estimate the consistency of the risk-reduction propositions and best consistency of voting the best shock-effect estimators prospectively (see Definition 1 and 2), which tell the probabilities that the prospective evaluation is consistent with the reality. Our Monte Carlo simulation results show that the risk-reduction conditions are highly consistent with the truth when the model for the shock effects is identified well with appropriate covariates under a fixed design; and gain more precision when the donor pool size increases. Our simulations further show that the procedure of voting the best technique by risk-reduction quantities is consistent with the truth when donor pool size is large. In the real data example of forecasting stock price of Conoco Phillips that experienced a shock on 2020 March 9th, the proposed three aggregation techniques work decently well. We now motivate our aggregation techniques.

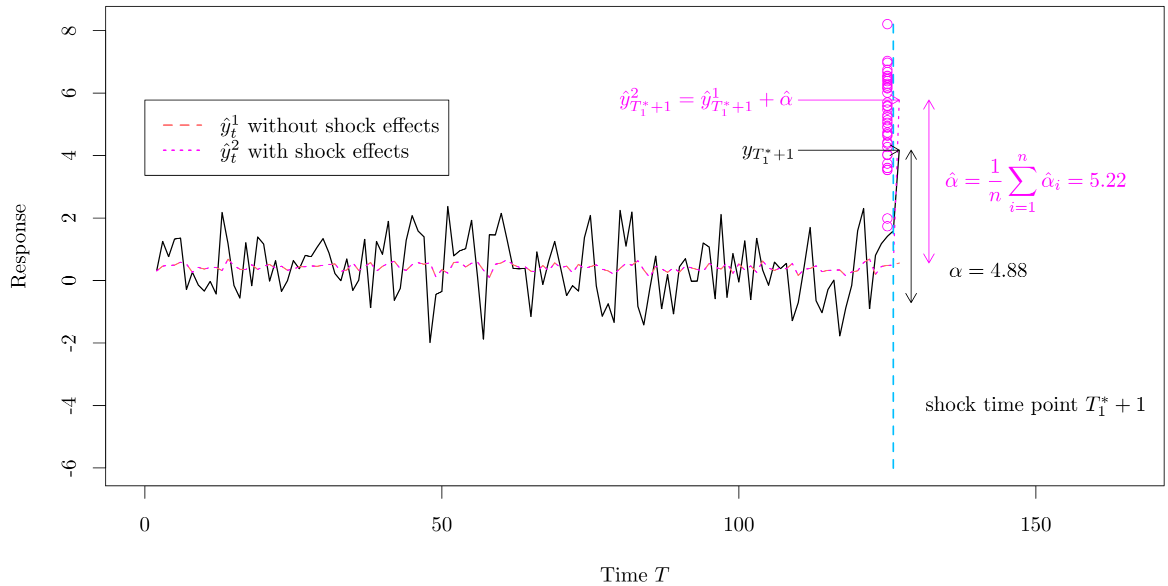

Figure 1. The time series experience a shock at $T_1^* + 1 = 126$ with true shock effect $\alpha = 4.88$. The figure is a comparison between forecast without considering shock effects and the one uses simple averaging given $n = 40$ disparate time series, and that the shock time is at $T_1^* + 1 = 126$. The magenta dots represent least square estimate $\hat{\alpha}_i$ from disparate time series. The prediction of $\hat{y}^2_{T_1^*+1}$ and $\hat{y}^1_{T_1^*+1}$ differs only by an adjustment $\hat{\alpha} = 5.22$. It is clear that $\hat{y}^2_{T_1^*+1}$ performs better than $\hat{y}^1_{T_1^*+1}$.

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.

We will suppose that specific interest is in forecasts to made after the occurrence of a structural shock. To gauge the performance of forecasts, we consider forecast risk in the form of mean squared error (MSE),

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

and root mean squared error (RMSE), given by $\sqrt{R_T}$, in our analyses.

Our post-shock prediction methodology will consist of selecting covariates $x_{i,t}$, constructing a suitable donor pool of candidate time series that have undergone similar structural shocks to the time series under study, and specifying a model for the time series $(y_{i,t}, \mathbf{x}_{i,t})$, $t = 1, \ldots, T_i$, $i = 1, \ldots, n+1$. 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 $D_{i,t}$, and we will assume that the donor pool time series are independent of the time series under study. 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 $T_1^* + 1 = 126$. It is supposed that the researcher does not have any information beyond $T_1^* + 1$, 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 $D_{i,t} = I(t = T_i^* + 1)$, $T_i^* < T_i$ 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:

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

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:

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

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

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

We further define $\tilde{\alpha}_i = \mu_\alpha + \delta_i' \mathbf{x}_{i,T_i^*+1} + \gamma_i' \mathbf{x}_{i,T_i^*}$. 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

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

where 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:

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

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

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

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

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

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

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

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

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

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

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

We further define

$$\mathbf{V} = \Big(\mathrm{vec}\big((\mathbf{x}_{2,T_2^*}, \mathbf{x}_{2,T_2^*+1})\big), \ldots, \mathrm{vec}\big((\mathbf{x}_{n+1,T_2^*}, \mathbf{x}_{n+1,T_2^*+1})\big) \Big).$$

Proposition 1. *If* $\mathbf{V}$ *has full rank and it exists some* $\mathbf{W}$ *satisfies (5), the solution to (6) is unique.*

Proposition 1 details some conditions when $\mathbf{W}^*$ is unique. Note that $\mathbf{V}$ is $2p \times n$. Therefore, if the covariates are of full rank and the true solution lies in the convex and compact $\mathcal{W}$, a sufficient condition for $\mathbf{W}^*$ to be unique is $2p \geq n$. However, when $2p < n$, $\mathbf{W}^*$ may not be unique. If it exists some $\mathbf{W}^*$ satisfies (5) and $2p < n$, there are infinitely many solutions to (5). The issue of non-uniqueness is further discussed in Section 6.

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 $\hat{\alpha}_{\text{adj}}$ and weighted-adjustment estimator. The expression for the expectations are as follow,

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

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

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

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

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

Formal justification for these results can be found in Appendix. Note that $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ are not unbiased estimators for α_1. Notice that under $\mathcal{M}_1$, $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{adj}}$ are unbiased estimators for $\mathrm{E}(\alpha_1) = \mu_\alpha$ (see distributional details of α_1 in Section 2.1). Nevertheless, $\hat{\alpha}_{\text{adj}}$ is a biased estimator for $\mathrm{E}(\alpha_1)$ but $\hat{\alpha}_{\text{wadj}}$ 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 2.

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

Unbiasedness properties for $\mathrm{E}(\alpha_1)$ 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

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

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

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

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

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

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

Formal justification for these results can be found in Appendix. Note that the variances are not comparable in closed-form because of the term $\mathrm{E}\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\}$. 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 2 implies that the adjustment estimator $\hat{\alpha}_{\mathrm{adj}}$ and weighted-adjustment estimator $\hat{\alpha}_{\mathrm{wadj}}$ 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 3. *Under $\mathcal{M}_1$,*

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

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

Proposition 3 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 3 shows that the risk-reduction condition is

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

In terms of the adjustment estimator, $\hat{\alpha}_{\mathrm{adj}}$, (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 $\hat{\alpha}_{\mathrm{wadj}}$, 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 3 and variance expression of $\hat{\alpha}_{\mathrm{wadj}}$, the following is the risk-reduction condition for $\hat{\alpha}_{\mathrm{wadj}}$.

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

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 $\hat{\alpha}_{\mathrm{wadj}}$. 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 2 implies that $\hat{\alpha}_{\text{wadj}}$ is an unbiased estimator of $\text{E}(\alpha_1)$. We now state conditions for risk reduction.

Proposition 4. *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 4, we can obtain a similar inequality as in Section 3.1.1 as below

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

where $\text{Var}(\alpha_i)$ may be replaced with 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 2 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 3 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 7.1) 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 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 $\Delta_{\text{wadj}} > \Delta_{\text{adj}}$ 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 3, 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 4 and the discussion in Section 3.1.2, the risk-reduction quantities $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$ are

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

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,

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

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 and risk-reduction quantities can be estimated using plug-in estimators in practice.

Remark 2. In Section 2.3, we noted that $\mathbf{W}^*$ may not be unique if $2p < n$. However, Proposition 3 and 4 will hold for every $\hat{\alpha}_{\text{wadj}}$ using $\mathbf{W}^*$ that satisfies (5).

3.3 Bootstrap for risk-reduction evaluation problems

In this section, we present a bootstrap procedure that approximates the distribution of our shock-effect estimators, checks the underlying conditions of our risk reduction propositions, and estimate risk-reduction using plug-in approach in practice. Our procedure involves the resampling of residuals in the separate OLS fits. This procedure has its origins in Section 6 of Efron and Tibshirani [1986], and it involves the resampling the residuals which are assumed to be the realizations of an iid processes. Bose [1988] showed that the asymptotic accuracy for OLS parameter estimation can be further improved from $O(T^{-1/2})$ to $o(T^{-1/2})$ almost surely under some regularity conditions.

To specify, there are two possible bootstrap procedures for our methods, one based on resampling the donor pool with replacement (*unconditional* on donor pool) and the other without any resampling of the donor pool (*conditional* on donor pool). We denote them by $\mathcal{B}_u$ and $\mathcal{B}_c$, respectively. We first illustrate the simple $\mathcal{B}_c$. The formal steps of it are outlined in the Supplementary Materials, the intuition for this procedure is as follows: let B be the bootstrap sample size and let $I = \{2, \ldots, n+1\}$ be the indexes for donor pool. Initialize $y_{i,0}$ for all $i \in I$. At iteration b, resample the residuals and then obtain shock-effect estimators for each of the disparate time series for all $i \in I$. Then construct and store any of the adjustment estimators $\hat{\alpha}_{\text{adj}}^{(b)}$, $\hat{\alpha}_{\text{wadj}}^{(b)}$, and $\hat{\alpha}_{\text{IVW}}^{(b)}$. We can then estimate distributional quantities of our shock-effect estimators with the bootstrap sample of $\hat{\alpha}_{\text{adj}}^{(b)}$, $\hat{\alpha}_{\text{wadj}}^{(b)}$, and $\hat{\alpha}_{\text{IVW}}^{(b)}$, for $b = 1, \ldots, B$. One can then use the bootstrapped shock-effects estimated by our residual bootstrap to provide an approximation for parameters involved in the risk-reduction conditions in Propositions 3 and 4. $\mathcal{B}_u$ is the same as above with the addition of first round of sampling n values from $I = \{2, \ldots, n+1\}$ without replacement to generate $I^{(b)}$. Then, it follows the same step with the one conditioned on the donor pool by replacing I by $I^{(b)}$.

Now, we discuss some consistency results for $\mathcal{B}_c$ conditioned on Θ_1. The bootstrap is aimed for estimating $\text{Var}(\hat{\alpha}_{\text{adj}})$, $\text{Var}(\hat{\alpha}_{\text{wadj}})$, and $\text{Var}(\hat{\alpha}_{\text{IVW}})$. The consistency results will be similar to the counterparts in sample mean. Therefore, we will briefly discuss the proof for the sample mean case and adapt it to the variance case. Let $\overline{\hat{\alpha}_{\text{adj}}} = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=2}^{n+1}\hat{\alpha}_i^{(b)}$, $\overline{\hat{\alpha}_{\text{wadj}}} = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=2}^{n+1} w_i^*\hat{\alpha}_i^{(b)}$, and $\overline{\hat{\alpha}_{\text{IVW}}} = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=2}^{n+1}\hat{\alpha}_{\text{IVW},i}^{(b)}$. The almost-sure convergence results for $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$ follow naturally, provided that the regularity conditions outlined in Bose [1988] hold. The proof for the case of $\overline{\hat{\alpha}_{\text{wadj}}}$, the sample mean of the bootstrapped weighted adjustment estimator, can be presented as below. As $B \to \infty$,

$$\overline{\hat{\alpha}_{\text{wadj}}} = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=2}^{n+1} w_i^*\hat{\alpha}_i^{(b)} = \sum_{i=2}^{n+1} w_i^*\frac{1}{B}\sum_{b=1}^{B}\hat{\alpha}_i^{(b)} \overset{a.s.}{\to} \sum_{i=2}^{n+1} w_i^*\alpha_i = \text{E}(\hat{\alpha}_{\text{wadj}}|\Theta).$$

The same rationale holds for the adjustment estimator as well since it is a linear combination of OLS estimates, and the weights are not random conditioned on Θ. However, the case for $\hat{\alpha}_{\text{IVW}}$ is slightly different because its weights are random. It is not clear with respect to whether similar consistency results holds for $\hat{\alpha}_{\text{IVW}}$. We would only claim that our bootstrap procedure provides an approximation for the case of $\hat{\alpha}_{\text{IVW}}$.

We stress that the above approximation is conditioned on Θ and for $\mathcal{B}_c$, and that bootstrapping cannot alleviate the inherent bias of using our adjustment estimators as estimates for α_1. For $\mathcal{B}_u$ but conditioned on Θ, we emphasize that the $\mathbf{W}^*$ from SCM method will become random, and that the consistency results may not hold. Simulation for justification of the parametric bootstrap is provided in Section 4.1.

Recall that $\hat{\alpha}_{\text{wadj}}$ and $\hat{\alpha}_{\text{IVW}}$ are unbiased estimators of their expectations, and $\hat{\alpha}_{\text{wadj}}$ is an unbiased estimator of $\text{E}(\alpha_1)$ from Proposition 2. Given the bootstrapped estimates of the variance, we can estimate the risk-reduction quantities in Section 3.2 using the estimates from the parametric bootstrap. For example, the Δ_{adj} and its estimator $\hat{\Delta}_{\text{adj}}$ are

$$\begin{aligned}
\Delta(\hat{\alpha}_{\text{adj}}) &= (\text{E}(\alpha_1))^2 - \text{Var}(\hat{\alpha}_{\text{adj}}) - (\text{E}(\hat{\alpha}_{\text{adj}}) - \text{E}(\alpha_1))^2 \\
\hat{\Delta}(\hat{\alpha}_{\text{adj}}) &= (\hat{\alpha}_{\text{wadj}})^2 - S^2_{\hat{\alpha}_{\text{adj}}} - (\hat{\alpha}_{\text{adj}} - \hat{\alpha}_{\text{wadj}})^2,
\end{aligned}$$

where $S^2_{\hat{\alpha}_{\text{adj}}}$ is the bootstrap sample variance for $\hat{\Delta}(\hat{\alpha}_{\text{adj}})$.

Suppose we have a pool of shock-effect estimators, $\mathcal{A}$, e.g., $\mathcal{A} = \{\hat{\alpha}_{\text{adj}}, \hat{\alpha}_{\text{wadj}}, \hat{\alpha}_{\text{IVW}}\}$ in our study. We can determine which one is the best shock-effect estimator. Suppose the $\hat{\Delta}(\hat{\alpha}_{\text{wadj}})$ and $\hat{\Delta}(\hat{\alpha}_{\text{IVW}})$ are estimators for $\Delta(\hat{\alpha}_{\text{wadj}})$ and $\Delta(\hat{\alpha}_{\text{IVW}})$, respectively. Then, we can evaluate the risk-reduction propositions by judging whether $\hat{\Delta}(\hat{\alpha}) > 0$ for $\hat{\alpha} \in \mathcal{A}$. Besides, we can also select the best shock-effect estimator accordingly. Define

$$\hat{\alpha}_{\text{best}} = \arg\max_{\hat{\alpha} \in \mathcal{A}} \hat{\Delta}(\hat{\alpha}) \quad \text{and} \quad \alpha_{\text{best}} = \arg\max_{\hat{\alpha} \in \mathcal{A}} \Delta(\hat{\alpha}). \tag{8}$$

That is, the best estimator $\hat{\alpha}_{\text{best}}$ is the one with the maximum estimated risk-reduction quantities whereas α_{best} is the true best shock-effect estimator given observed post-shock response. See more discussions in Section 3.4.

In terms of the use of our bootstrap procedure, we caution that the bootstrapping residuals in OLS estimation may not provide valid inference in moderate or high dimension where $p < T_i$ but p/T_i is not close to zero for $i \in \{2, \ldots, n+1\}$ [El Karoui and Purdom, 2018]; see alternatives for residual bootstrapping in linear models in El Karoui and Purdom [2018].

Recall that $\mathbf{W}^*$ may not be unique if the conditions in Proposition 1 are not satisfied. Under the setup of our model, non-uniqueness of $\mathbf{W}^*$ would not be a problem for inferential purposes. It is because all risk-reduction propositions and other properties established will still hold. However, non-uniqueness may not be desirable in other model setup. For example, consider the model where α_i is assumed to be not identically distributed, the size of donor pool to be 2, and there are two solutions to (5), say, $\mathbf{W}_1^* = (1, 0)$ and $\mathbf{W}_2^* = (0, 1)$ with $\text{Var}(\alpha_2) \neq \text{Var}(\alpha_3)$. In this scenario, the procedure of trying to recover α_1 will fail since the resulting α_1 will be different with different variances for $\mathbf{W}_1^* = (1, 0)$ and $\mathbf{W}_2^* = (0, 1)$. If the non-uniqueness is of concern, users may select the weight that optimizes some objective function. For instance, in the example just discussed, users may select the weight that minimizes the estimated variance of $\hat{\alpha}_{\text{wadj}}$.

There are some issues in using $\mathcal{B}_u$ when $\mathbf{W}^*$ falls in the boundary of the parameter space. See detailed discussions in Section 6.

3.4 Leave-one-out cross validation with k random draws

In this section, we introduce a powerful procedure for rendering prospective evaluations of the applicability of the proposed methods. First, we introduce the concept of consistency. Second, we detail the leave-one-out cross-validation with k random draws to realize the estimation prospectively.

Our risk-reduction propositions methodologies can be evaluated by *consistency*. Risk-reduction propositions are decision-making procedures that might commit a type of error, which is similar to Type I or Type II error in a hypothesis test. Let $\hat{\alpha}$ be an estimator of α_1 and $\delta_{\hat{\alpha}}$ is the corresponding risk-reduction proposition. If $\Delta(\hat{\alpha}) > 0$ ($\Delta(\hat{\alpha}) < 0$, respectively) but $\delta_{\hat{\alpha}}$ incorrectly reported $\Delta(\hat{\alpha}) < 0$ ($\Delta(\hat{\alpha}) > 0$, respectively) so that it make the decision not to use $\hat{\alpha}$ (to use $\hat{\alpha}$, respectively), $\delta_{\hat{\alpha}}$ is said to be *inconsistent*. If $\Delta(\hat{\alpha}) < 0$ ($\Delta(\hat{\alpha}) > 0$, respectively) and $\delta_{\hat{\alpha}}$ correctly reported $\Delta(\hat{\alpha}) < 0$ ($\Delta(\hat{\alpha}) > 0$, respectively) so that it make the decision to use $\hat{\alpha}$, $\delta_{\hat{\alpha}}$ is said to be *consistent*. These situations are depicted in the following table.

		Decision	
		Use $\hat{\alpha}$	Do not use $\hat{\alpha}$
Truth	$\Delta(\hat{\alpha}) > 0$	Consistent	Inconsistent
	$\Delta(\hat{\alpha}) < 0$	Inconsistent	Consistent

Similar definitions can be adapted to the best shock-effect estimators.

In this setting, judging whether δ_α is consistent can be treated as a Bernoulli trial that reports 1 if δ_α is consistent, and reports 0 if δ_α is inconsistent. It is also true for selecting the best shock-effect estimator. It motivates us to define

$$\mathcal{C}(\delta_\alpha) = I(\delta_{\hat{\alpha}} \text{ is consistent}) \quad \text{and} \quad \mathcal{C}(\mathcal{A}) = I(\hat{\alpha}_{\text{best}} = \alpha_{\text{best}})$$

where $I(\cdot)$ is an indicator function, $\mathcal{A}$ is the pool of shock-effect estimators, α_{best} is the true best shock-effect estimator, and $\hat{\alpha}_{\text{best}}$ is the best shock-effect estimator selected by the maximum estimated risk-reduction, see Section 3.3. Then, we can define consistency and best consistency as in Definition 1 and 2.

Definition 1. The *consistency* of the risk-reduction proposition $\delta_{\hat{\alpha}}$ is $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$.

Definition 2. Given $\mathcal{A}$ the pool of shock-effect estimators, the *best consistency* is $\mathrm{E}(\mathcal{C}(\mathcal{A}))$.

Consistency and best consistency are important parameters for users to evaluate the effectiveness of our risk-reduction propositions and estimated risk-reduction. If $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}})) > 0.5$, we claim that $\delta_{\hat{\alpha}}$ is better than a random guess. If $\mathrm{E}(\mathcal{C}(\mathcal{A})) > 1/|\mathcal{A}|$, we claim that the procedure of selecting the best shock-effect estimators is better than a random guess. Note that $\mathcal{C}(\delta_{\hat{\alpha}})$ and $\mathcal{C}(\mathcal{A})$ generally can be known only if the post-shock response is observed. But it is only feasible in a retrospective finding. It is even more difficult to construct a random sample to estimate $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ or $\mathrm{E}(\mathcal{C}(\mathcal{A}))$. Nevertheless, it is possible to estimate them using the procedure *leave-one-out cross validation (LOOCV)*.

Our proposed LOOCV tailors the standard LOOCV to our application. See details of standard LOOCV in Section 7.10 of Hastie et al. [2009]. Recall in Section 2.1 that we are given the data $\{(\mathbf{x}_{i,t}, y_{i,t})\colon i = 1,\dots,n+1, t = 1,\dots,T_i\}$, where $\{(\mathbf{x}_{1,t}, y_{1,t})\colon t = 1,\dots,T_1\}$ is the data of the time series of interest and the remaining is those of the donor pool. For m^{th} iteration of LOOCV, where $m = \{1,\dots,n\}$, we set aside $\{(\mathbf{x}_{m+1,t}, y_{m+1,t})\colon t = 1,\dots,T_{m+1}\}$ as the time series of interest, and construct a new donor pool $\{(\mathbf{x}_{i,t}, y_{i,t})\colon i \in \mathcal{I}, t = 1,\dots,T_i\}$, where $\mathcal{I} = \{2,\dots,n+1\} \setminus \{m+1\}$. Since the post-shock response $y_{m+1,T^*_{m+1}+1}$ is observed, for $\hat{\alpha} \in \mathcal{A}$, we can compute the consistency and best consistency as $\mathcal{C}^{(-m)}(\delta_{\hat{\alpha}})$ and $\mathcal{C}^{(-m)}(\mathcal{A})$ using the estimation procedures in Section 3.3. The the LOOCV estimates for $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ and $\mathrm{E}(\mathcal{C}(\mathcal{A}))$ are

$$\bar{\mathcal{C}}(\delta_{\hat{\alpha}}) = \frac{1}{n}\sum_{m=1}^{n} \mathcal{C}^{(-m)}(\delta_{\hat{\alpha}}) \quad \text{and} \quad \bar{\mathcal{C}}(\mathcal{A}) = \frac{1}{n}\sum_{m=1}^{n} \mathcal{C}^{(-m)}(\mathcal{A}).$$

Note that to satisfy $\mathcal{M}_2$ in Section 2.1, the candidates from donor pool has to be assumed to be mutually independent. Based on this assumption, $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$ and $\bar{\mathcal{C}}(\mathcal{A})$ should be almost unbiased estimates of $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ and $\mathrm{E}(\mathcal{C}(\mathcal{A}))$, respectively [Marden, 2015, Page 222].

However, LOOCV can be very computationally intensive when n is large. In an analog to k-fold cross-validation as remedies for ordinary LOOCV, we propose a similar procedure *LOOCV with k random draws* to alleviate the computation burden. LOOCV with k random draws differs in the selection of indices. The algorithms are outlined as follows. If $n \le k$, we set k to be n. If $n > k$, we randomly sample without replacement k elements from $\{1,\dots,n\}$ to gather them into $\mathcal{J}$. For $m \in \mathcal{J}$, we set aside $\{(\mathbf{x}_{m+1,t}, y_{m+1,t})\colon t = 1,\dots,T_{m+1}\}$ as the time series of interest, and construct a new donor pool $\{(\mathbf{x}_{i,t}, y_{i,t})\colon i \in \mathcal{I}, t = 1,\dots,T_i\}$, where $\mathcal{I} = \{2,\dots,n+1\} \setminus \{m+1\}$. The LOOCV with k random draws estimates for $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ and $\mathrm{E}(\mathcal{C}(\mathcal{A}))$ are

$$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}}) = \frac{1}{k}\sum_{m\in\mathcal{J}} \mathcal{C}^{(-m)}(\delta_{\hat{\alpha}}) \quad \text{and} \quad \bar{\mathcal{C}}^{(k)}(\mathcal{A}) = \frac{1}{k}\sum_{m\in\mathcal{J}} \mathcal{C}^{(-m)}(\mathcal{A}).$$

Similarly, assuming the candidates in the donor pool are mutually independent and the data satisfy $\mathcal{M}_2$, $\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}})$ and $\bar{\mathcal{C}}^{(k)}(\mathcal{A})$ should be almost unbiased estimates of $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ and $\mathrm{E}(\mathcal{C}(\mathcal{A}))$, respectively. The estimation bias should decrease as $k \to n$. In practice, $k = 5$ or $k = 10$ can be typical choices for LOOCV with k random draws.

4 Simulation

In this section, we provide justification for our methods based on Monte Carlo simulation examples. We implemented our simulation based on $\mathcal{M}_{22}$ with negligibly small Σ_γ and Σ_δ approximating the design of

Table 1: 30 Monte Carlo simulations for $\mathcal{B}_c$ with varying n and σ_α (risk-reduction propositions $\delta_{\hat{\alpha}}$)

n	σ_α	$\hat{\alpha}_{\text{adj}}$	$\hat{\alpha}_{\text{wadj}}$	$\hat{\alpha}_{\text{IVW}}$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{adj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{wadj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{IVW}}})$	$\bar{\mathcal{C}}^{(k)}(\mathcal{A})$
	5	1 (0)	1 (0)	1 (0)	0.89 (0.03)	0.90 (0.02)	0.89 (0.03)	0.35 (0.04)
	10	1 (0)	1 (0)	1 (0)	0.89 (0.03)	0.89 (0.03)	0.89 (0.03)	0.37 (0.05)
5	25	1 (0)	1 (0)	1 (0)	0.78 (0.03)	0.83 (0.03)	0.78 (0.03)	0.44 (0.05)
	50	0.90 (0.06)	0.97 (0.03)	0.93 (0.05)	0.66 (0.04)	0.65 (0.04)	0.65 (0.04)	0.45 (0.06)
	100	0.77 (0.08)	0.97 (0.03)	0.80 (0.07)	0.57 (0.04)	0.53 (0.04)	0.57 (0.04)	0.49 (0.05)
	5	1 (0)	1 (0)	1 (0)	0.91 (0.02)	0.92 (0.02)	0.91 (0.02)	0.25 (0.03)
	10	1 (0)	1 (0)	1 (0)	0.89 (0.02)	0.89 (0.03)	0.89 (0.02)	0.29 (0.03)
10	25	1 (0)	1 (0)	1 (0)	0.75 (0.03)	0.78 (0.03)	0.77 (0.04)	0.3 (0.04)
	50	0.83 (0.07)	1 (0)	0.80 (0.07)	0.59 (0.04)	0.63 (0.04)	0.59 (0.04)	0.37 (0.05)
	100	0.80 (0.07)	0.93 (0.05)	0.80 (0.07)	0.47 (0.04)	0.51 (0.04)	0.46 (0.04)	0.41 (0.05)
	5	1 (0)	1 (0)	1 (0)	0.91 (0.02)	0.93 (0.02)	0.91 (0.02)	0.31 (0.05)
	10	1 (0)	1 (0)	1 (0)	0.87 (0.02)	0.89 (0.03)	0.87 (0.02)	0.31 (0.04)
15	25	1 (0)	1 (0)	1 (0)	0.75 (0.03)	0.78 (0.03)	0.76 (0.03)	0.37 (0.04)
	50	0.933 (0.05)	1 (0)	0.9 (0.06)	0.61 (0.04)	0.69 (0.03)	0.64 (0.04)	0.41 (0.04)
	100	0.67 (0.09)	1 (0)	0.63 (0.09)	0.55 (0.04)	0.51 (0.04)	0.55 (0.04)	0.47 (0.04)
	5	1 (0)	1 (0)	1 (0)	0.95 (0.02)	0.94 (0.02)	0.95 (0.02)	0.29 (0.04)
	10	1 (0)	1 (0)	1 (0)	0.93 (0.02)	0.91 (0.02)	0.93 (0.02)	0.30 (0.04)
25	25	1 (0)	1 (0)	1 (0)	0.78 (0.04)	0.79 (0.04)	0.77 (0.04)	0.31 (0.04)
	50	0.90 (0.06)	1 (0)	0.9 (0.06)	0.57 (0.04)	0.60 (0.04)	0.58 (0.04)	0.34 (0.04)
	100	0.83 (0.07)	1 (0)	0.80 (0.07)	0.49 (0.04)	0.48 (0.04)	0.50 (0.04)	0.39 (0.03)

$\mathcal{M}_{21}$. We consider $p = 13$ and $\mu_\alpha = 2$, where $p = 13$ is set to satisfy conditions in Proposition 1. Parameter setup of our simulations is detailed as below. ϕ_i's are sampled independently from Uniform$(0, 1)$. We sampled T_i's independently from $\Gamma(15, 10)$ that are further rounded to integers. If it exists some i such that $T_i < 90$, we let $T_i = 90$ instead. Moreover, T_i^* is randomly selected from $2p + 4$ to $T_i - 1$. Those are set up to satisfy a necessary condition for the design matrix of OLS estimation to have full rank. Moreover, it is also designed to illustrate the performance of $\hat{\alpha}_{\text{IVW}}$ that may perform well in time series with varying lengths. Additionally, we generated the covariates from $\Gamma(1, 2)$ to set up a setting when the $\hat{\alpha}_{\text{wadj}}$ may perform well. Last, we set $\gamma_i, \delta_i \overset{iid}{\sim} \mathcal{N}(1, 0.5)$ and $\theta_i, \beta_i \sim \mathcal{N}(0, 1)$.

In this experiment, we consider parameter setup by varying σ in the model of $y_{i,t}$, n, the donor pool size, and σ_α in the model of α_i. We used 30 replications of Monte Carlo simulations and $B = 200$ for computation; and report means and standard errors for estimated quantities accordingly. $k = 5$ is set up for LOOCV with k random draws.

4.1 Bootstrap simulation for risk-reduction propositions

In this section, we discuss simulation results for the bootstrap procedures used in estimating parameters for risk-reduction propositions (see Section 3) and risk-reduction quantities (see Section 3.2). We will compare $\mathcal{B}_c$ with $\mathcal{B}_u$.

In the first experiment, we consider the parameter combination of $n \in \{5, 10, 15, 25\}$ and $\sigma_\alpha \in \{5, 10, 25, 50, 100\}$ where we fix σ to 10. Note that $\mathrm{E}(\mathrm{E}(\alpha_1)) = 54$, where the last expectation is operated under the density of the covariates. In other words, data with $\sigma_\alpha \in \{5, 10, 25, 50, 100\}$ should well represent the situations when the signal of the covariates is strong and when it is nearly lost.

Table 1 shows the results, with the 3rd to 5th column representing the averaged $I\big(\hat{\Delta}(\hat{\alpha}) > 0\big)$ (i.e.,

Table 2: 30 Monte Carlo simulations for $\mathcal{B}_c$ with varying σ and σ_α

σ	σ_α	$\hat{\alpha}_{\text{adj}}$	$\hat{\alpha}_{\text{wadj}}$	$\hat{\alpha}_{\text{IVW}}$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{adj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{wadj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{IVW}}})$	$\bar{\mathcal{C}}^{(k)}(\mathcal{A})$
	5	1 (0)	1 (0)	1 (0)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	0.4 (0.04)
	10	1 (0)	1 (0)	1 (0)	0.91 (0.02)	0.95 (0.02)	0.92 (0.02)	0.46 (0.04)
5	25	0.97 (0.03)	1 (0)	0.97 (0.03)	0.83 (0.03)	0.89 (0.02)	0.83 (0.03)	0.43 (0.04)
	50	0.87 (0.06)	1 (0)	0.87 (0.06)	0.67 (0.04)	0.69 (0.03)	0.67 (0.04)	0.38 (0.04)
	100	0.7 (0.09)	0.93 (0.05)	0.77 (0.08)	0.52 (0.04)	0.51 (0.04)	0.51 (0.05)	0.51 (0.05)
	5	1 (0)	1 (0)	1 (0)	0.93 (0.02)	0.93 (0.02)	0.93 (0.02)	0.27 (0.03)
	10	1 (0)	1 (0)	1 (0)	0.86 (0.03)	0.89 (0.03)	0.87 (0.03)	0.37 (0.04)
10	25	0.97 (0.03)	1 (0)	0.97 (0.03)	0.78 (0.03)	0.82 (0.03)	0.79 (0.03)	0.29 (0.04)
	50	0.93 (0.05)	1 (0)	0.9 (0.06)	0.66 (0.04)	0.67 (0.04)	0.66 (0.04)	0.32 (0.04)
	100	0.77 (0.08)	0.97 (0.03)	0.7 (0.09)	0.54 (0.05)	0.54 (0.04)	0.55 (0.04)	0.43 (0.05)
	5	0.97 (0.03)	1 (0)	1 (0)	0.78 (0.03)	0.77 (0.03)	0.79 (0.03)	0.22 (0.03)
	10	1 (0)	1 (0)	1 (0)	0.81 (0.03)	0.79 (0.03)	0.81 (0.03)	0.21 (0.03)
25	25	1 (0)	1 (0)	1 (0)	0.72 (0.03)	0.77 (0.03)	0.71 (0.03)	0.25 (0.04)
	50	0.87 (0.06)	0.97 (0.03)	0.87 (0.06)	0.55 (0.04)	0.57 (0.04)	0.52 (0.04)	0.35 (0.04)
	100	0.9 (0.06)	1 (0)	0.87 (0.06)	0.54 (0.04)	0.54 (0.04)	0.56 (0.04)	0.39 (0.05)
	5	0.83 (0.07)	0.73 (0.08)	0.8 (0.07)	0.54 (0.04)	0.57 (0.04)	0.55 (0.04)	0.19 (0.03)
	10	0.77 (0.08)	0.8 (0.07)	0.77 (0.08)	0.53 (0.04)	0.55 (0.05)	0.53 (0.04)	0.23 (0.03)
50	25	0.73 (0.08)	0.8 (0.07)	0.73 (0.08)	0.58 (0.04)	0.58 (0.04)	0.57 (0.04)	0.21 (0.03)
	50	0.8 (0.07)	0.83 (0.07)	0.8 (0.07)	0.59 (0.04)	0.53 (0.04)	0.6 (0.04)	0.29 (0.05)
	100	0.5 (0.09)	0.73 (0.08)	0.53 (0.09)	0.51 (0.05)	0.48 (0.05)	0.51 (0.05)	0.31 (0.04)
	5	0.43 (0.09)	0.37 (0.09)	0.43 (0.09)	0.47 (0.04)	0.47 (0.04)	0.47 (0.04)	0.19 (0.03)
	10	0.63 (0.09)	0.6 (0.09)	0.67 (0.09)	0.51 (0.04)	0.49 (0.04)	0.51 (0.04)	0.23 (0.03)
100	25	0.57 (0.09)	0.6 (0.09)	0.57 (0.09)	0.51 (0.04)	0.53 (0.04)	0.49 (0.04)	0.21 (0.04)
	50	0.53 (0.09)	0.43 (0.09)	0.5 (0.09)	0.47 (0.04)	0.49 (0.05)	0.49 (0.05)	0.19 (0.03)
	100	0.63 (0.09)	0.63 (0.09)	0.63 (0.09)	0.43 (0.04)	0.41 (0.04)	0.45 (0.04)	0.17 (0.03)

guess of $\delta_{\hat{\alpha}}$), and the 6th to 9th columns representing the LOOCV consistency and best consistency following notations in Section 3.4. First, assuming that $\bar{C}^{(k)}(\delta_{\hat{\alpha}})$ well estimates $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ and fixing n, we observe that $\delta_{\hat{\alpha}}$ is highly consistent for $\hat{\alpha} \in \mathcal{A}$ when σ_α is small from Table 1. The reasons can be explained as follows. When σ_α is small, the signal of the covariates is strong so that $\hat{\alpha}_{\text{wadj}}$ will be expected to capture the signal according to construction of $\hat{\alpha}_{\text{wadj}}$ in Section 2.3. Moreover, when σ_α is small, $\mathcal{M}_{22}$ approximates $\mathcal{M}_{21}$ such that estimation of $\mathrm{E}(\alpha_1)$ should be nearly unbiased according to Proposition 2. However, when the signal of the covariates is poor (σ_α is big), the $\delta_{\hat{\alpha}}$ become more inconsistent for $\hat{\alpha} \in \mathcal{A}$. It is to be expected since the bootstrap estimates become more biased. However, users can be warned by $\bar{C}^{(k)}(\delta_{\hat{\alpha}})$ to have an idea of the effectiveness of $\delta_{\hat{\alpha}}$. Second, fixing σ_α, we can observe that $\delta_{\hat{\alpha}}$ becomes more consistent when n increases. It is due to the robustness gain in estimation when n increases.

Interestingly, we observe different patterns of $\bar{\mathcal{C}}^{(k)}(\mathcal{A})$ as opposed to $\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}})$: (1) as n increases, $\bar{\mathcal{C}}^{(k)}(\mathcal{A})$ decreases; and (2) as σ_α increases, $\bar{\mathcal{C}}^{(k)}(\mathcal{A})$ increases. It is possible when n is small or σ_α is large, the differences among $\hat{\Delta}(\hat{\alpha})$ for $\hat{\alpha} \in \mathcal{A}$ are large so that the ranking gains more precision. It is because in these situations, the estimation of bootstrap estimates is less precise so that the variance of $\hat{\Delta}(\hat{\alpha})$ is larger.

Additionally, we observe that in most cases $\delta_{\hat{\alpha}_{\text{wadj}}}$ reports $\hat{\alpha}_{\text{wadj}}$ reduces the risk even when $\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{wadj}}})$ starts to break down. Recall from Section 3.3 that $\hat{\Delta}(\hat{\alpha})$ contains the squared bias for estimating $\mathrm{E}(\alpha_1)$. But it is not present for $\hat{\Delta}(\hat{\alpha}_{\text{wadj}})$ since we applied the fact $\hat{\alpha}_{\text{wadj}}$ is unbiased for $\mathrm{E}(\alpha_1)$ from Proposition

Table 3: 30 Monte Carlo simulation with varying n and σ_α (distance to y_{1,T_1^*+1})

n	σ_α	Original	$\hat{\alpha}_{\text{adj}}$	$\hat{\alpha}_{\text{wadj}}$	$\hat{\alpha}_{\text{IVW}}$
	5	51.92 (4.04)	19.23 (2.55)	20.64 (2.76)	19.36 (2.52)
	10	52.58 (4.35)	21.565 (2.58)	23.38 (2.86)	21.72 (2.50)
5	25	55.02 (5.84)	30.30 (3.59)	31.99 (4.36)	30.20 (3.50)
	50	64.42 (8.11)	50.00 (5.78)	52.56 (6.92)	49.55 (5.70)
	100	91.51 (13.20)	94.27 (10.41)	97.26 (12.57)	93.61 (10.27)
	5	52.34 (4.00)	17.23 (2.96)	18.55 (2.84)	17.39 (2.95)
	10	52.59 (4.05)	19.02 (3.24)	20.95 (3.14)	19.23 (3.24)
10	25	54.07 (4.97)	27.66 (4.55)	31.79 (4.47)	27.85 (4.61)
	50	60.32 (7.53)	47.70 (7.04)	52.97 (7.54)	47.78 (7.17)
	100	85.60 (12.99)	89.86 (12.82)	100.74 (13.61)	90.40 (12.91)
	5	49.86 (4.01)	18.07 (2.88)	18.38 (2.71)	18.04 (2.88)
	10	48.73 (4.30)	19.45 (2.97)	19.35 (2.86)	19.32 (3.00)
15	25	47.06 (5.13)	26.16 (3.31)	26.81 (3.13)	26.23 (3.33)
	50	48.75 (6.86)	40.27 (4.43)	42.09 (4.49)	40.77 (4.38)
	100	64.29 (11.11)	68.85 (8.27)	74.08 (8.72)	69.91 (8.14)
	5	57.60 (6.94)	21.58 (5.90)	20.00 (5.86)	21.58 (5.90)
	10	56.80 (7.01)	22.20 (5.97)	20.47 (5.89)	22.22 (5.97)
25	25	56.58 (7.49)	28.96 (6.49)	27.06 (6.19)	29.03 (6.48)
	50	64.33 (8.75)	47.16 (8.28)	46.30 (7.26)	47.35 (8.26)
	100	95.61 (13.02)	90.10 (13.29)	86.81 (11.75)	90.55 (13.23)

2. Therefore, when the signal from covariates is poorer, $\delta_{\hat{\alpha}_{\text{wadj}}}$ becomes less conservative. Besides, the averaged $I\big(\hat{\Delta}(\hat{\alpha}) > 0\big)$ times $\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}})$ can provide an approximation for the probability that $\hat{\alpha}$ actually reduces the risk assuming an symmetry of consistencies between the cases when $\hat{\Delta}(\hat{\alpha}) > 0$ and when $\hat{\Delta}(\hat{\alpha}) < 0$. For example, when $n = 5$ and $\sigma_\alpha = 50$, the probability that $\hat{\alpha}_{\text{adj}}$ reduces the risk is approximately $0.90 \times 0.66 = 0.594$ from table 1. In other words, the probability that $\hat{\alpha}$ reduces the risk has the same pattern as $\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}})$ has with n and σ_α for $\hat{\alpha} \in \mathcal{A}$.

In the second experiment, we consider the parameter combination of $\sigma, \sigma_\alpha \in \{5, 10, 25, 50, 100\}$ where we fix n to 10. Likewise, $\sigma, \sigma_\alpha \in \{5, 10, 25, 50, 100\}$ will produce situations when the signal of the covariates is strong and when it is nearly lost in the model of both $y_{i,t}$ and α_i.

From Table 2, we observe that as σ_α increases fixing σ, $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$ and $\bar{\mathcal{C}}(\mathcal{A})$ will decrease, which is a pattern similar to one shown in the first experiment. Furthermore, as σ increases fixing σ_α, $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$ and $\bar{\mathcal{C}}(\mathcal{A})$ decrease as well. Note that the consistency or best consistency hinges on the estimation of the parameters. Since $\hat{\alpha}_{\text{wadj}}$ is a linear combination of OLS estimates, as σ increases, $\text{Var}(\hat{\alpha}_{\text{wadj}})$ increases as well. Therefore, $\hat{\alpha}_{\text{wadj}}$ become more volatile and its estimation of $\text{E}(\alpha_1)$ can be less reliable. Those reasons can explain why an increase of σ_α contributes to a decrease of $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$ and $\bar{\mathcal{C}}(\mathcal{A})$.

With respect to averaged $I(\hat{\Delta}(\hat{\alpha} > 0)$ (i.e., the guess), it starts to decrease as σ increases. It is reasonable if we believe the bootstrap estimate $S^2_{\hat{\alpha}}$ provides a good approximation for $\text{Var}(\hat{\alpha})$ for $\hat{\alpha} \in \mathcal{A}$. The reasons can be outlined as below. Recall in Section 3.1.2, the conditions of risk-reduction propositions involve $(\text{E}(\alpha_1))^2 > \text{Var}(\hat{\alpha}) + (\text{E}(\hat{\alpha}) - \text{E}(\alpha_1))^2$ for $\hat{\alpha} \in \mathcal{A}$, where we note that those parameters are *true* ones but not estimated ones. Notice that $\text{Var}(\hat{\alpha})$ is a function of σ since $\hat{\alpha}$ is estimated by OLS. Therefore, it explains the reason why the increase of σ would result in a decrease of averaged $I(\hat{\Delta}(\hat{\alpha} > 0)$ since the conditions are not likely to hold when $\text{Var}(\hat{\alpha})$ increases.

Simulation for $\mathcal{B}_u$ with the same parameter setup as that of $\mathcal{B}_c$ are implemented. See table 6 and

Table 4: 30 Monte Carlo simulation with varying n and σ_α (distance to y_{1,T_1^*+1})

σ	σ_α	Original	$\hat{\alpha}_{\text{adj}}$	$\hat{\alpha}_{\text{wadj}}$	$\hat{\alpha}_{\text{IVW}}$
	5	53.27 (2.59)	10.91 (1.70)	11.77 (1.55)	10.8 (1.69)
	10	50.86 (3.83)	18.06 (2.18)	17.26 (2.17)	18.25 (2.17)
5	25	60.52 (3.97)	16.09 (2.40)	19.83 (3.05)	15.87 (2.38)
	50	54.65 (6.95)	51.79 (6.87)	53.83 (7.26)	52.6 (6.91)
	100	104.73 (13.00)	88.31 (12.16)	88.88 (12.63)	86.72 (12.35)
	5	58.17 (4.18)	18.17 (2.61)	16.59 (2.39)	18.09 (2.63)
	10	52.81 (4.07)	19.05 (2.49)	19.67 (2.66)	18.95 (2.53)
10	25	61.53 (5.69)	28.55 (4.07)	31.82 (4.27)	28.76 (4.07)
	50	56.31 (8.03)	47.33 (4.89)	41.44 (4.39)	46.86 (4.79)
	100	84.38 (11.62)	82.91 (11.77)	84.5 (12.39)	84.3 (11.93)
	5	56.35 (6.1)	25.54 (4.06)	30.49 (4.61)	25.94 (4.02)
	10	49.8 (5.24)	25.04 (4.38)	26.74 (3.64)	24.86 (4.38)
25	25	54.21 (6.51)	44.17 (6.66)	43.41 (6.95)	43.89 (6.63)
	50	66.51 (7.86)	46.19 (8.04)	44.83 (9.31)	46.54 (8.27)
	100	109.21 (13.29)	78.47 (12.45)	83.99 (12.00)	78.57 (12.75)
	5	75.29 (10.75)	63.51 (8.30)	64.25 (8.88)	63.32 (8.40)
	10	57.59 (6.00)	48.51 (8.32)	50.58 (8.63)	48.08 (8.36)
50	25	77.21 (12.03)	54.73 (9.47)	54.76 (10.24)	54.2 (9.62)
	50	90.48 (10.21)	68.88 (9.20)	68.28 (10.48)	68.8 (9.28)
	100	111.09 (17.54)	102.47 (16.68)	110.15 (15.87)	101.53 (16.52)
	5	214.07 (67.40)	195.46 (67.65)	197.51 (68.45)	196.58 (67.51)
	10	120.85 (15.39)	114.79 (15.51)	119.19 (15.75)	114.02 (15.74)
100	25	97.84 (13.75)	95.02 (14.88)	100.04 (16.94)	96.23 (14.71)
	50	141.22 (24.51)	136.11 (25.73)	150.62 (27.62)	135.26 (25.39)
	100	95.53 (12.80)	103.01 (14.12)	103.31 (16.27)	105.79 (14.32)

table 7 for results. Comparing table 1 and table 6 yields that when n is small ($n = 5$ or $n = 10$) and σ_α is small ($\sigma_\alpha = 5$), $\mathcal{B}_u$ is better than $\mathcal{B}_c$ with statistical evidence. For other situations, $\mathcal{B}_u$ and $\mathcal{B}_c$ are rather similar. It is likely that the extra randomness from sampling with replacement from donor pool compensates for the possible noises from a small donor pool. Concerning table 2 and table 7, it appears that when $n = 10$ and $\sigma_\alpha = 5$, $\mathcal{B}_u$ is better than $\mathcal{B}_c$ when σ increases. It might be the case that additional layer of bootstrap in the donor pool buffers the negative effects on $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$ introduced from increasing variation of $y_{i,t}$. However, when σ_α increases over 5 and $n = 10$, $\mathcal{B}_c$ and $\mathcal{B}_u$ are quite similar under situations of different σ and σ_α. In conclusion, $\mathcal{B}_u$ is better than $\mathcal{B}_c$ when the signal of the covariates is strong and n is small; otherwise, they are similar. See more discussions for differences between $\mathcal{B}_u$ and $\mathcal{B}_c$ in Section 6.

4.2 Simulation for prediction

In this section, we report prediction results for our methods in the simulation. Specifically, we are interested in the distance between the true response y_{1,T_1^*+1} and the adjusted forecasts $\hat{y}_{1,T_1^*+1}$, $\hat{y}_{1,T_1^*+1} + \hat{\alpha}_{\text{adj}}$, $\hat{y}_{1,T_1^*+1} + \hat{\alpha}_{\text{wadj}}$, and $\hat{y}_{1,T_1^*+1} + \hat{\alpha}_{\text{IVW}}$.

In the first experiment, table 3 shows that similar to Section 4.1, as σ_α increases, the prediction appears to be poorer. When $\sigma_\alpha = 5, 10, 25$, forecasts using $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ are always better than the original forecast significantly. But it does not hold generally for the case when $\sigma_\alpha = 50, 100$. It is reasonable in that when the σ_α is large, it is difficult to find a reliable estimate of α_1. Nevertheless, unlike what is detected in Section 4.1, no statistical evidence has been found to support the claim that n matters in prediction. In other words, n matters in producing reliable decision-making of $\delta_{\hat{\alpha}}$ rather than reliable prediction.

Note that there is no statistical evidence supporting for differences among $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$. This observation corresponds to the fact that $\bar{\mathcal{C}}(\mathcal{A})$ is small and following analysis in Section 4.1. When their performances do not differ much, the precision of ranking is small. In this scenario, for a more reasonable statistical inference, one may opt for a double bootstrap instead to approximate the distribution of $\Delta(\hat{\alpha})$ and compare estimators by checking whether their 95% confidence intervals overlap. See more details in Section 6.

In the second experiment, table 4 shows similar results for prediction to those in the first expeirment, for the case of varying σ_α but fixing σ, and the case of varying σ but fixing σ_α. When σ increases with fixing σ_α, it is likely that the degree of variation of $y_{1,t}$ exceeds the extent of adjustment improvement $\hat{\alpha}$ can contribute to for $\hat{\alpha} \in \mathcal{A}$. It can be confirmed by looking into the case of $\sigma = 50, 100$ as opposed to the case for $\sigma = 5, 10, 25$ where adjusted forecasts outperform original forecasts. Adjusted forecasts can be worse than original forecast if the variation of $y_{i,t}$ is large and the signal of the covariates is very small. See for the case of $\sigma = 100$ and $\sigma_\alpha = 100$ in table 4.

5 Forecasting Conoco Phillips stock in the presence of shocks

In this example we forecast Conoco Phillips stock prices in the midst of the coronavirus recession. Specific interest is in predictions made after March 6, 2020, the Friday before the stock market crash on March 9, 2020. We will detail how we combine knowledge from disparate time series to improve the forecast of Conoco Phillips stock price that would be made without adjustments for the shock.

Conoco Phillips is chosen for this analysis because it is a large oil and gas resources company [ConocoPhillips, 2020]. Focus on the oil sector is because oil prices have been shown to exhibit a cointegrating behavior with economic indices [He et al., 2010], and our chosen time frame represents the onset of a significant economic down turn, coupled with a Russia and OPEC battle for global oil price control the Sunday before trading resumes on Monday, March 9th [Sukhankin, 2020]. Furthermore, fear of and action in response to the coronavirus pandemic began to uptick dramatically between Friday, March 6th and Monday, March 9th. Major events include the SXSW festival being cancelled as trading closed on March 6th [Wang et al., 2020]. New York declared a state of emergency on March 7th [New York State Government, 2020], and by Sunday, March 8th, eight states have declared a state of emergency [Alonso, 2020] while Italy placed 16 million people in quarantine [Sjödin et al., 2020].

Economic indicators forecasted our recession before the coronavirus pandemic began. The current recession followed an inversion of the yield curve that first happened back in March, 2019 [Tokic, 2019]. An inversion of the yield curve is an event that signals that recessions are more likely [Andolfatto and Spewak, 2018, Bauer and Mertens, 2018]. In this analysis we investigate the performance of oil companies in previous recessions that followed an inversion of the yield curve to obtain a suitable Conoco Phillips donor pool for estimating the March 9th shock effect on Conoco Phillips oil stock. We also consider previous OPEC oil supply shocks [Mensi et al., 2014]. We will borrow from the literature on oil price forecasting to establish appropriate time horizons and forecasting models. Recessions that occurred before 1973 are disregarded since oil price forecasts cannot be represented by standard time series models before 1973 [Alquist et al., 2013]. In this analysis we make the following considerations:

(1) **AR(1) model and time window**. We will use an AR(1) model to forecast Conoco Phillips stock price. This model has been shown to beat no-change forecasts when predicting oil prices over time horizons of one and three months [Alquist et al., 2013]. We will consider 30 pre-shock trading days and we will forecast the immediate shock effect and the shock effect over a future five trading day

window. All estimates will be adjusted for inflation. The model setup for AR(1) is exactly the same as what is stated in Section 2.1 with addition of shock effects. All the parameters are estimated using OLS.

(2) **Selection of covariates**. We perform our analyses incorporating daily S&P 500 index prices and West Texas Intermediate (WTI) crude oil prices as covariates.

(3) **Construction of donor pool**. Our donor pool will consist of Conoco Phillips shock effects observed on March 14, 2008, several events in September, 2008, and November 27, 2014. The first two shock effects were observed during recessions that were predicated by an inversion of the yield curve [Bauer and Mertens, 2018], and the third was an OPEC induced supply side shock effect [Huppmann and Holz, 2015]. The reasons for those three shocks are:

   (a) On March 14, 2008, Bear Stearns was verging on bankruptcy from what its officials described as a sudden liquidity squeeze related to its large exposure to devalued mortgage-backed securities. On that day, it also received word that it was getting an unprecedented loan from the Federal Reserve System, this decision was unprecedented: never before had the Fed committed to "bailing out" a financial entity that was not a commercial bank. The day of the announcement, the stocks of other major Wall Street firms tumbled (including Conoco Phillips). These concerns then spilled over into the broader universe of stocks [Shorter, 2008].

   (b) In early September 2008, time series of oil prices experienced a sudden increase in volatility simultaneously due to turmoil in financial markets. The political, economic, social or environmental events may coincide with these shocks [Ewing and Malik, 2013]. Notable shock effects followed the placement of Fannie May and Freddie Mac in conservatorship on September 7th (shock effect on the 8th), Lehman Brothers filing for bankruptcy on September 15th, and the Office of Thrift Supervision closes Washington Mutual Bank on September 25th [Dwyer and Tkac, 2009, Longstaff, 2010].

   (c) On November 27th, 2014, it is documented that oil prices fall as OPEC opts not to cut production [Huppmann and Holz, 2015]. During the Great Recession when economic activity clearly declined, both oil and stock prices fell which points to demand factors. During the second half of 2014, oil prices plummeted but equity prices generally increased, suggesting that supply factors were the key driver [Baffes et al., 2015, Page 19].

We assume that the five shocks are independent of the shock that Conoco Phillips experienced on March 9, 2020. The covariates and response of time series in the donor pool are adjusted for inflation. Note that there are three shock-effects nested in the time series 2008 September, we assume that these three shocks are independent, where the assumption checks using likelihood ratio test are provided in Appendix 7.2. We computed 3 shock-effect estimates. Using the same model that is described in $\mathcal{M}_2$ (see Section 2.1), we computed 3 shock-effect estimates, namely the adjustment $\hat{\alpha}_{\text{adj}}$, weighted adjustment $\hat{\alpha}_{\text{wadj}}$, and the inverse-variance weighted $\hat{\alpha}_{\text{IVW}}$. For $\hat{\alpha}_{\text{wadj}}$, the fitting result can be described by

$$\mathbf{W}^* = (0.000, 0.000, 0.000, 0.000, 1.000) \quad \text{and} \quad \left\| \text{vec}\big(\mathbf{X}_1 - \hat{\mathbf{X}}_1(\mathbf{W}^*)\big) \right\|_4 = 1314.04.$$

Note that the norm is computed using the k-dimensional Euclidean metric. The solution $\mathbf{W}^*$ reports that the α_1 is very similar to α_5 (November 27, 2014) under the setup of $\mathcal{M}_2$ and our collected covariates.

Table 5: Bootstrap estimates and results yielded by risk-reduction propositions with $B = 1000$

	$\hat{\alpha}_{\text{adj}}$	$\hat{\alpha}_{\text{wadj}}$	$\hat{\alpha}_{\text{IVW}}$
Bootstrapped variance	0.531	0.479	0.970

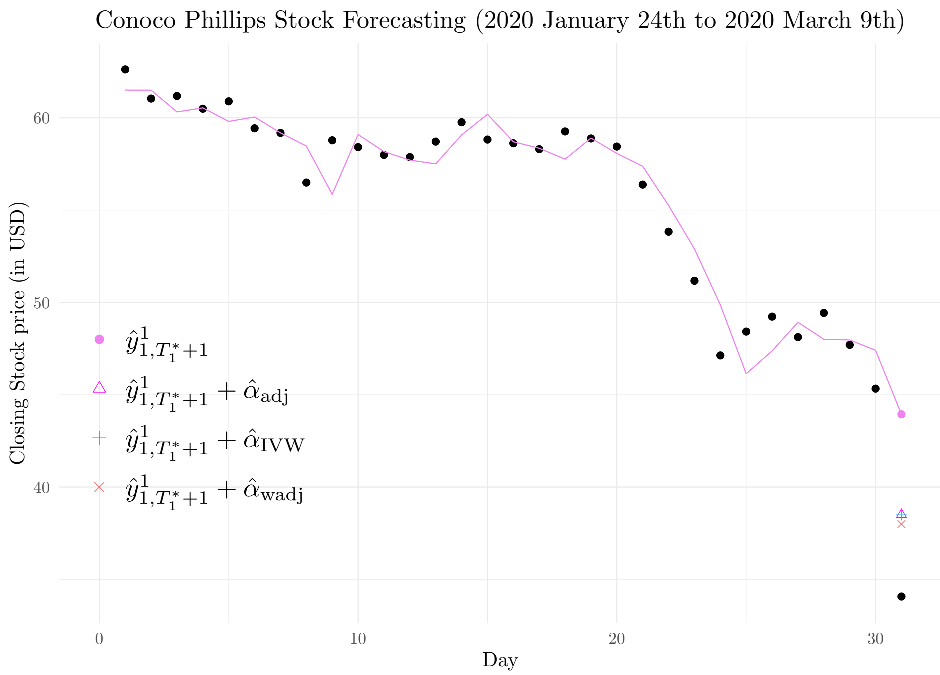

Figure 2: March 9th, 2020 post-shock forecasts for Conoco Phillips stock price.

Using $\mathcal{B}_c$ proposed in Section 3.3, we estimated parameters for risk-reduction propositions and risk-reduction quantities proposed in Section 3; and the result is presented in Table 2. Plugging these estimates into conditions in Section 3 and risk-reduction formulas in Section 3.2 yields (1) $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ reduce the risk and (2) the risk-reduction quantities are 34.565, 34.880, and 34.139, respectively. The estimated risk-reduction quantities vote $\hat{\alpha}_{\text{wadj}}$ for the best estimator. LOOCV with k random draws was not implemented since the effective n is 3, a number too small to provide a credible estimation. We verify the consistency of the result yielded by risk-reduction propositions with the reality as below.

We can see from Figure 2 that $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$ and $\hat{\alpha}_{\text{IVW}}$ perform decently well, they do not recover the magnitude of the shock effect but are much better than unadjusted forecasts that do not account for shock effects. The unadjusted forecast has an RMSE of 9.870 dollars whereas the use of $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ have RMSE of 4.436, 3.924, and 4.423 dollars, respectively. Therefore, the risk-reduction propositions are consistent with the reality with the reduced risks for forecasts using $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ than without. Moreover, the risk-reduction quantities are consistent with the reality as well.

The phenomenon that the true shock effect is not recovered by $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ can be due to that the donor pool is not constructed to be similar enough to the time series of interest. The shock(s) on March 9, 2020 is(are) in the midst of the COVID-19 pandemic and oil production volatility. It is difficult to find available stock market time series data that were generated under a similar setting.

From another perspective, it is possible that the stock of Conoco Phillips actually experienced multiple shocks on 2020 March 9th. For example, Kilian [2009] studied the effect that different supply and demand shocks have on oil prices through a vector auto regressive model. Their model postulates an additive nature of shock effects, although the additivity parameters requires estimation in their context. Motivated by his study, we also studied additive shock effect estimators where the shock effects corresponding to separate supply and demand shocks are added to estimate the unknown shock effect. The supply shock donor pool consists of the November 27, 2014 shock effect; and the demand shock donor pool consists of the remaining shock effects. The additive adjustment estimator computed by adding the $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ estimators for the demand and supply shock effects have RMSEs of 1.382, 2.468, and 1.145 dollars, respectively. These additive adjustment estimators do extremely well, nearly perfectly forecasting the realized shock effect.

6 Discussion

Our proposed model in Section 2.1 is a simple model with results in Section 3 easily generalizing to other more complicated models. Our simple AR(1) model can be extended to AR(p) settings. Moreover, the mean function for α_i under $\mathcal{M}_2$ can be extended to include more lagged predictors. Similarly, LOOCV with k random draws can be adapted to AR(p). Besides, the functional form for the mean function for $y_{i,t}$ can be extended beyond the linear regression model under $\mathcal{M}_2$. Multiple shock-effects can be nested within a time series; and time series in the donor pool are allowed be dependent. As an example, we could consider a dependency structure for the September 2008 shock effects in our analysis of Conoco Phillips stock. But we note that consistency estimates from LOOCV with k random draws may not work well if donor pool candidates are not mutually independent since the almost unbiased property hinges on the mutual independence among candidates in the donor pool. Although it is reflected in $\mathcal{M}_2$, we stress that our proposed methods allow α_i to follow arbitrary distributions with existing first and second moments. The covariates in the model for α_i under $\mathcal{M}_2$ can be different from the covariates in the model of $y_{i,t}$. We also note that our post-shock framework can be extended to settings where the shock effect can be decomposed into separable estimable parts. An example of this is the additive shock effect estimators that we studied in our Conoco Phillips analysis.

Our bootstrap procedures can be extended to approximate the distribution of shock effect estimators from more general time series. The pseudo time series generated by our proposed parametric bootstrap is not stationary. Note that Politis and Romano [1994] motivated a stationary bootstrap method for strictly stationary and weakly dependent time series $\{X_n \colon n \in \mathbb{N}\}$. This algorithm generates a sequence of blocks of observations $B_{I_1,L_1}, B_{I_2,L_2}, \ldots$ where $B_{i,b} = \{X_i, X_{i+1}, \ldots, X_{i+b-1}\}$; for $j > N$, X_j is defined to be X_k, where $k = j \mod N$ and $X_0 = X_N$. The sampling stops when N observations are reached. Note that the collection of random positions $\{I_n \colon n \in \mathbb{N}\}$ is a sequence of i.i.d. discrete uniform random variables; and the collection of random lengths $\{L_n \colon n \in \mathbb{N}\}$ is a sequence of i.i.d. geometric random variables with parameter p. However, the consistency needs to be proved by a case-by-case analysis [Politis et al., 1999, Page 66]. Additionally, the asymptotic accuracy of this algorithm can be sensitive to the selection of p. 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 Chapters 3 and 4 in Politis et al. [1999], and Berkowitz and Kilian [2000]. It is up to users in terms of selecting which procedure to choose but under different assumptions on the time series.

Recall that in Section 3.3, we proposed two possible bootstrap procedures, $\mathcal{B}_u$ and $\mathcal{B}_c$. There are some philosophical distinctions between $\mathcal{B}_u$ and $\mathcal{B}_c$. $\mathcal{B}_u$ treats the donor pool as realizations from some infinite super-population of potential donors. In contrast, $\mathcal{B}_c$ treats the donor pool as being fixed and known before the analysis is conducted, where the randomness comes from parameters and idiosyncratic error. For other differences, Section 4.1 shows that $\mathcal{B}_u$ is better than $\mathcal{B}_c$ when the signal of the covariates is strong and n is small; otherwise, they are quite similar.

In this study, for the use of $\mathcal{B}_u$, we assume $\mathbf{W}^*$ is non-degenerate in the population without taking zero weight. Recall that in Section 2.3 we note that if there exists some $\mathbf{W}^*$ satisfies (5) and $2p < n$, there will be infinitely many solutions to $\mathbf{W}^*$. In the simulation setup, we consider $2p > n$ such that the assumption can be reasonably satisfied. However, in the application, it is likely that $\mathbf{W}^*$ may take values on the boundary of $\mathcal{W}$, in which case $\mathcal{B}_u$ may fail for $\hat{\alpha}_{\text{wadj}}$ [Andrews, 2000]. For example, if $\mathbf{W}^*$ is observed to fall in the boundary, $\mathcal{B}_u$ will fail. Moreover, if $2p < n$, $\mathcal{B}_u$ fails since the existence of infinitely many solutions is certain (if there exists some $\mathbf{W}^*$ satisfies (5)), and will guarantee degeneracy of $\mathbf{W}^*$. In this regard, the subsampling in Section 4 of Li [2019] may adapt to our scenario to solve this problem by resampling m candidates from donor pool with the conditions $m \to \infty$, and $m/n \to 0$ as $n \to \infty$. But we may leave its mathematical justification for future research. Note that this problem will not occur under $\mathcal{B}_c$ because if $w_i^* = 0$ for some i and the weights are known, $w_i^* = 0$ makes it impossible for bootstrap estimate of $\hat{\alpha}_i$ to contribute to the bootstrap estimate of distribution of $\hat{\alpha}_{\text{wadj}}$.

Double bootstrap can be another general bootstrap alternative to the one proposed in Section 3.3.

For example, for $\mathcal{B}_u$, double bootstrap adds another layer of residual bootstrap between the donor pool non-parametric bootstrap and the original residual bootstrap. It enables an estimation for distribution of $\Delta(\hat{\alpha})$ for $\hat{\alpha} \in \mathcal{A}$. In the framework of double bootstrap, instead of checking whether $\Delta(\hat{\alpha}) > 0$, we can check whether non-parametric 95% confidence intervals contain 0. Likewise, we can compare two shock-effect estimators by checking whether their 95% confidence intervals of risk-reduction overlap though an exhaustive ranking may not be feasible. We have attempted Monte Carlo simulations using the same setting in Section 4 for the double bootstrap procedure, the results of which are quite similar to those of Section 3.3.

For the use of our methodology, caution should be dedicated to the construction of donor pool. If the donor pool includes some individuals that are not similar to the time series of interest, the result will possibly not be robust to the introduced noises. Moreover, according to Section 4.1, we recommend using estimated risk-reduction quantities for voting best shock-effect estimator only when n is large. It is due to the fact that plug-in estimation of risk-reduction quantities in Section 3.3 gains more precision when n becomes large.

7 Appendix

7.1 Proofs

7.1.1 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

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

Similarly,

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

7.1.2 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

$$\mathrm{Var}(\hat{\alpha}_{\text{wadj}}) = \mathrm{E}(\mathrm{Var}(\hat{\alpha}_{\text{wadj}}|\Theta)) + \mathrm{Var}(\mathrm{E}(\hat{\alpha}_{\text{wadj}}|\Theta))$$

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

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

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

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

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

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

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

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

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

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 $\mathcal{M}_{21}$.

7.1.3 Proofs for lemmas and propositions

Proof of Proposition 1 The proof of Li [2019] in Appendix A.2 and A.3 adapts easily to Proposition 1. ☐

Proof of Proposition 2 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.

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

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

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

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

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,

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

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 3 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 2, the property that $\hat{\alpha}_{\text{wadj}}$ 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 4 By Proposition 2, the property that $\hat{\alpha}_{\text{wadj}}$ is an unbiased estimator of $\mathrm{E}(\alpha_1)$ holds for $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$. The remainder of the proof follows a similar argument to the proof of Proposition 3. □

7.2 Supplementary materials for data analysis

The independence of estimated shock-effects are further tested using likelihood ratio test (LRT) based on their estimated covariance matrix. The estimated covariance matrix is

$$\hat{\mathbf{\Sigma}} = \begin{pmatrix} 4.012 & 0.362 & -0.062 \\ 0.362 & 3.894 & -0.029 \\ -0.062 & -0.029 & 3.927 \end{pmatrix}.$$

with degrees of freedoms 35. Using LRT for independence between blocks of random variables [Marden, 2015, Section 10.2], the LRT test statistic is 0.304 with p-value of 0.581. Therefore, we do not reject the null hypothesis that the three estimated shock-effects are independent.

7.3 Simulations tables for $\mathcal{B}_u$

Table 6: 30 Monte Carlo simulations for $\mathcal{B}_u$ with varying n and σ_α

		Guess			LOOCV with k random draws			
n	σ_α	$\hat{\alpha}_{\text{adj}}$	$\hat{\alpha}_{\text{wadj}}$	$\hat{\alpha}_{\text{IVW}}$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{adj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{wadj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{IVW}}})$	$\bar{\mathcal{C}}^{(k)}(\mathcal{A})$
	5	1 (0)	1 (0)	1 (0)	0.92 (0.02)	0.96 (0.02)	0.92 (0.02)	0.36 (0.05)
	10	1 (0)	1 (0)	1 (0)	0.90 (0.02)	0.92 (0.02)	0.90 (0.02)	0.40 (0.04)
5	25	0.97 (0.03)	1 (0)	0.97 (0.03)	0.80 (0.02)	0.81 (0.02)	0.80 (0.03)	0.43 (0.04)
	50	0.83 (0.07)	0.87 (0.06)	0.87 (0.06)	0.55 (0.05)	0.57 (0.05)	0.55 (0.05)	0.38 (0.05)
	100	0.47 (0.09)	0.73 (0.08)	0.47 (0.09)	0.48 (0.05)	0.48 (0.04)	0.46 (0.04)	0.37 (0.05)
	5	1 (0)	1 (0)	1 (0)	0.95 (0.02)	0.95 (0.02)	0.95 (0.02)	0.33 (0.04)
	10	1 (0)	1 (0)	1 (0)	0.92 (0.03)	0.91 (0.03)	0.92 (0.03)	0.33 (0.04)
10	25	0.90 (0.06)	0.97 (0.03)	0.93 (0.05)	0.77 (0.04)	0.79 (0.04)	0.75 (0.05)	0.31 (0.04)
	50	0.77 (0.08)	0.80 (0.07)	0.77 (0.08)	0.55 (0.04)	0.64 (0.04)	0.55 (0.05)	0.31 (0.04)
	100	0.63 (0.09)	0.70 (0.09)	0.63 (0.09)	0.53 (0.04)	0.53 (0.05)	0.51 (0.04)	0.33 (0.04)
	5	1 (0)	1 (0)	1 (0)	0.92 (0.02)	0.93 (0.02)	0.92 (0.02)	0.31 (0.05)
	10	1 (0)	1 (0)	1 (0)	0.92 (0.02)	0.91 (0.02)	0.92 (0.02)	0.30 (0.03)
15	25	1 (0)	1 (0)	1 (0)	0.83 (0.03)	0.83 (0.03)	0.83 (0.03)	0.35 (0.04)
	50	0.90 (0.06)	0.93 (0.05)	0.90 (0.06)	0.71 (0.03)	0.67 (0.04)	0.69 (0.04)	0.33 (0.04)
	100	0.70 (0.09)	0.73 (0.08)	0.70 (0.09)	0.54 (0.05)	0.61 (0.04)	0.54 (0.05)	0.31 (0.05)
	5	1 (0)	1 (0)	1 (0)	0.90 (0.03)	0.91 (0.03)	0.90 (0.03)	0.27 (0.04)
	10	1 (0)	1 (0)	1 (0)	0.87 (0.03)	0.89 (0.03)	0.87 (0.03)	0.30 (0.03)
25	25	1 (0)	1 (0)	1 (0)	0.73 (0.03)	0.74 (0.04)	0.73 (0.03)	0.29 (0.03)
	50	0.83 (0.07)	0.87 (0.06)	0.83 (0.07)	0.55 (0.05)	0.56 (0.05)	0.55 (0.05)	0.29 (0.04)
	100	0.77 (0.08)	0.73 (0.08)	0.77 (0.08)	0.48 (0.05)	0.49 (0.04)	0.49 (0.05)	0.31 (0.05)

Table 7: 30 Monte Carlo simulations for $\mathcal{B}_u$ with varying σ and σ_α

		Guess			LOOCV with k random draws			
σ	σ_α	$\hat{\alpha}_{\text{adj}}$	$\hat{\alpha}_{\text{wadj}}$	$\hat{\alpha}_{\text{IVW}}$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{adj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{wadj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{IVW}}})$	$\bar{\mathcal{C}}^{(k)}(\mathcal{A})$
	5	1 (0)	1 (0)	1 (0)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	0.39 (0.05)
	10	1 (0)	1 (0)	1 (0)	0.97 (0.01)	0.97 (0.01)	0.97 (0.01)	0.34 (0.04)
5	25	0.93 (0.05)	0.97 (0.03)	0.93 (0.05)	0.81 (0.04)	0.82 (0.04)	0.81 (0.04)	0.29 (0.04)
	50	0.77 (0.08)	0.80 (0.07)	0.77 (0.08)	0.57 (0.05)	0.67 (0.04)	0.57 (0.05)	0.34 (0.04)
	100	0.63 (0.09)	0.70 (0.09)	0.63 (0.09)	0.51 (0.04)	0.53 (0.05)	0.50 (0.04)	0.35 (0.05)
	5	1 (0)	1 (0)	1 (0)	0.95 (0.02)	0.95 (0.02)	0.95 (0.02)	0.33 (0.04)
	10	1 (0)	1 (0)	1 (0)	0.92 (0.03)	0.91 (0.03)	0.92 (0.03)	0.33 (0.04)
10	25	0.90 (0.06)	0.97 (0.03)	0.93 (0.05)	0.77 (0.04)	0.79 (0.04)	0.75 (0.05)	0.31 (0.04)
	50	0.77 (0.08)	0.80 (0.07)	0.77 (0.08)	0.55 (0.04)	0.64 (0.04)	0.55 (0.05)	0.31 (0.04)
	100	0.63 (0.09)	0.70 (0.09)	0.63 (0.09)	0.53 (0.04)	0.53 (0.05)	0.51 (0.04)	0.33 (0.04)
	5	1 (0)	1 (0)	1 (0)	0.79 (0.04)	0.78 (0.04)	0.77 (0.04)	0.19 (0.04)
	10	0.97 (0.03)	0.97 (0.03)	0.97 (0.03)	0.75 (0.04)	0.75 (0.04)	0.73 (0.05)	0.20 (0.04)
25	25	0.90 (0.06)	0.93 (0.05)	0.90 (0.06)	0.66 (0.05)	0.69 (0.05)	0.66 (0.05)	0.30 (0.05)
	50	0.77 (0.08)	0.80 (0.07)	0.80 (0.07)	0.55 (0.05)	0.64 (0.05)	0.55 (0.05)	0.34 (0.05)
	100	0.60 (0.09)	0.70 (0.09)	0.63 (0.09)	0.48 (0.04)	0.52 (0.04)	0.47 (0.04)	0.31 (0.04)
	5	0.73 (0.08)	0.73 (0.08)	0.77 (0.08)	0.57 (0.05)	0.57 (0.05)	0.58 (0.05)	0.24 (0.04)
	10	0.73 (0.08)	0.73 (0.08)	0.73 (0.08)	0.55 (0.05)	0.55 (0.05)	0.55 (0.05)	0.23 (0.03)
50	25	0.77 (0.08)	0.70 (0.09)	0.73 (0.08)	0.48 (0.05)	0.51 (0.05)	0.47 (0.05)	0.24 (0.04)
	50	0.77 (0.08)	0.73 (0.08)	0.73 (0.08)	0.46 (0.05)	0.53 (0.04)	0.47 (0.05)	0.25 (0.04)
	100	0.57 (0.09)	0.70 (0.09)	0.57 (0.09)	0.46 (0.05)	0.49 (0.03)	0.45 (0.05)	0.33 (0.05)
	5	0.63 (0.09)	0.60 (0.09)	0.63 (0.09)	0.46 (0.05)	0.52 (0.04)	0.48 (0.05)	0.23 (0.04)
	10	0.60 (0.09)	0.60 (0.09)	0.63 (0.09)	0.46 (0.05)	0.53 (0.04)	0.46 (0.05)	0.24 (0.04)
100	25	0.63 (0.09)	0.57 (0.09)	0.63 (0.09)	0.45 (0.05)	0.54 (0.04)	0.47 (0.05)	0.21 (0.04)
	50	0.57 (0.09)	0.53 (0.09)	0.57 (0.09)	0.49 (0.04)	0.54 (0.04)	0.47 (0.05)	0.23 (0.03)
	100	0.43 (0.09)	0.53 (0.09)	0.43 (0.09)	0.52 (0.05)	0.55 (0.03)	0.51 (0.05)	0.27 (0.04)

References

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

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

Melissa Alonso. At least 8 us states have declared a state of emergency. https://www.cnn.com/asia/live-news/coronavirus-outbreak-03-08-20-intl-hnk/h_1b09bcd8c4b247c893d65b7118353923, 2020. Accessed on 2020-06-01.

Ron Alquist, Lutz Kilian, and Robert J Vigfusson. Forecasting the price of oil. In *Handbook of economic forecasting*, volume 2, pages 427–507. Elsevier, 2013.

David Andolfatto and Andrew Spewak. Does the yield curve really forecast recession? *Economic Synopses*, (30):1–2, 2018.

Donald WK Andrews. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68(2):399–405, 2000.

John Baffes, M Ayhan Kose, Franziska Ohnsorge, and Marc Stocker. The great plunge in oil prices: Causes, consequences, and policy responses. *Consequences, and Policy Responses (June 2015)*, 2015.

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.

Michael D. Bauer and Thomas M. Mertens. Economic forecasts with the yield curve. *Federal Reserve Bank of San Francisco Economic Letter*, pages 1–5, 2018.

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.

ConocoPhillips. What we do. http://www.conocophillips.com/about-us/how-energy-works/, 2020. Accessed on 2020-05-24.

Gerald P Dwyer and Paula Tkac. The financial crisis of 2008 in fixed-income markets. *Journal of International Money and Finance*, 28(8):1293–1316, 2009.

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

Jana Eklund and Sune Karlsson. Forecast combination and model averaging using predictive measures. *Econometric Reviews*, 26(2-4):329–363, 2007.

Noureddine El Karoui and Elizabeth Purdom. Can we trust the bootstrap in high-dimensions? the case of linear models. *The Journal of Machine Learning Research*, 19(1):170–235, 2018.

Bradley T Ewing and Farooq Malik. Volatility transmission between gold and oil futures under structural breaks. *International Review of Economics & Finance*, 25:113–121, 2013.

Yue Fang. Forecasting combination and encompassing tests. *International Journal of Forecasting*, 19(1): 87–94, 2003.

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.

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

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

Yanan He, Shouyang Wang, and Kin Keung Lai. Global economic activity and crude oil prices: A cointegration analysis. *Energy Economics*, 32(4):868–876, 2010.

David F Hendry and Michael P Clements. Pooling of forecasts. *The Econometrics Journal*, 7(1):1–31, 2004.

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

Daniel Huppmann and Franziska Holz. What about the opec cartel? Technical report, DIW Roundup: Politik im Fokus, 2015.

Lutz Kilian. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–69, 2009.

Gary Koop and Simon Potter. Forecasting in dynamic factor models using bayesian model averaging. *The Econometrics Journal*, 7(2):550–565, 2004.

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 C.B. Phillips. Model selection in the presence of incidental parameters. *Journal of Econometrics*, 188(2):474–489, 2015.

Jiahan Li and Weiye Chen. Forecasting macroeconomic time series: Lasso-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting*, 30(4):996–1015, 2014.

Kathleen T Li. Statistical inference for average treatment effects estimated by synthetic control methods. *Journal of the American Statistical Association*, pages 1–16, 2019.

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

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

Francis A Longstaff. The subprime credit crisis and contagion in financial markets. *Journal of financial economics*, 97(3):436–450, 2010.

John I Marden. Multivariate statistics: Old school. *University of Illinois*, 2015.

Walid Mensi, Shawkat Hammoudeh, and Seong-Min Yoon. How do opec news and structural breaks impact returns and volatility in crude oil markets? further evidence from a long memory process. *Energy Economics*, 42:343–354, 2014.

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

New York State Government. At novel coronavirus briefing, governor cuomo declares state of emergency to contain spread of virus. https://www.governor.ny.gov/news/novel-coronavirus-briefing-governor-cuomo-declares-state-emergency-contain-spread-virus, 2020. Accessed on 2020-05-24.

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

M Hashem Pesaran and Andreas Pick. Forecast combination across estimation windows. *Journal of Business & Economic Statistics*, 29(2):307–318, 2011.

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.

Dimitris N Politis, Joseph P Romano, and Michael Wolf. *Subsampling*. Springer Science & Business Media, 1999.

Gary W Shorter. *Bear Stearns: Crisis and" rescue" for a major provider of mortgage-related products*. Congressional Research Service, 2008.

Henrik Sjödin, Annelies Wilder-Smith, Sarah Osman, Zia Farooq, and Joacim Rocklöv. Only strict quarantine measures can curb the coronavirus disease (covid-19) outbreak in italy, 2020. *Eurosurveillance*, 25(13):2000280, 2020.

Sergey Sukhankin. Russian geopolitical objectives in the current oil price crisis, and implications for canada. *The School of Public Policy Publications*, 13, 2020.

Norman R Swanson and Halbert White. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, 79(4): 540–550, 1997.

Norman R Swanson and Tian Zeng. Choosing among competing econometric forecasts: Regression-based forecast combination using model selection. *Journal of Forecasting*, 20(6):425–440, 2001.

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

Damir Tokic. Is yield curve predicting a us recession in 2020? *Journal of Corporate Accounting & Finance*, 30(3):5–7, 2019.

Xutong Wang, Remy F Pasco, Zhanwei Du, Michaela Petty, Spencer J Fox, Alison P Galvani, Michael Pignone, S Claiborne Johnston, and Lauren Ancel Meyers. Impact of social distancing measures on covid-19 healthcare demand in central texas. *medRxiv*, 2020.

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.