Minimizing post-shock forecasting error using disparate information

Jilei Lin* and Daniel J. Eck†

Department of Statistics, University of Illinois at Urbana-Champaign

July 21, 2020

Abstract

We developed a forecasting methodology for providing credible forecasts for the time series data that have 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 forecasts 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. Leave-one-out cross validation is proposed to estimate correctness of risk-reduction propositions, prospectively informing users of the probabilities that this prospective evaluation is in line 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. The core idea of our methodology is to sensibly aggregate similar past realized shock-effects which arose from disparate time-series, and then incorporate the aggregated shock-effect estimator into the present forecast. Our method of combining disparate shock effects embraces ideas from time series pooling using cross-sectional panel data (??????), judgement forecasting (??), synthetic control methodology (??), and expectation shocks (??). We study this methodology in the context of additive shock-effects in linear autoregressive models.

In our post-shock forecasting setting, the researcher has a time series of interest which is known to have recently undergone a structural shock, and the post-shock response is not observed. The researcher must therefore move beyond the modeling paradigm that they were previously working under to accommodate this new shock, as in judgement forecasting (??). In our methodological framework, the researcher creates a synthetic panel of disparate time series which have undergone similar structural shocks in the past. Construction of the donor pool that forms this synthetic panel is similar to that in synthetic control methodology (SCM) (?). As in SCM, care is needed when forming the donor pool of disparate time series. However, there are key differences between our framework and SCM. We assume that the disparate time series are independent from the time series under study before the timing of the shock. We also assume that the shock-effects for each disparate time series are independent realizations from some unknown distribution.

*jileil2@ilinois.edu

†dje13@illinois.edu

We estimate the shock-effects that are present in the disparate time series for which post-shock responses are observed. We then aggregate these estimated shock-effects and use this aggregated estimate as an estimator for the shock-effect in the time series of interest. This estimator is then added to a forecast for the yet to be realized post-shock response corresponding to the time series of interest. Shock-effects in our post-shock forecasting framework is similar to “expectation shocks” which are studied in **?**. The context in **?** allowed for consistent estimation of expectation shocks under a vector autoregressive model, possibly involving an instrumental variable approach as in **?**. We offer no such consistency guarantee here. In our context, the yet to observed shock-effect of interest is a realization from a random process that can only be partially explained by the realizations of the disparate time series.

In this article, we will assume a simple auto regressive data generating process similar to that in **?** with a general random effects structure. Therefore, our methodology is similar to the “K latent pooling” framework of **?**. However, our model formulation is more general than **?** and our donor pool is formed from independent time series. 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 (**?**). These adjustment strategies all target the mean of the shock-effect distribution. However, such an estimation strategy can reduce mean squared error (MSE) when variation in the shock-effect distribution is small relative to the mean. We provide risk-reduction propositions that detail the conditions when the adjusted forecasts will work better than the original forecast. The involved parameters in the risk-reduction propositions and risk-reduction quantities can be estimated by a residual bootstrap procedure that we develop within. We also motivate a simple leave-one-out cross validation procedure which can prospectively assess the performance of our shock-effect adjustment estimators. This prospective assessment does not require the observation of the post-shock response. Our Monte Carlo simulation results show that the risk-reduction propositions are nearly perfectly correct when the model for the shock effects is identified well with appropriate covariates under a fixed design. We demonstrate the utility of our methodology in a real data analysis in which we forecast the stock price of Conoco Phillips shares that experienced a large structural shock on March 9th, 2020. We show that our proposed adjustment estimators yield much better results than no adjustment in this setting. We also use this example to demonstrate settings in which the shock-effect may be decomposed into separate estimable parts. We now motivate our framework for post-shock forecasting.

2 Setting

We will suppose that a researcher 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 forecasting the response 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 forecasting methodology will consist of selecting covariates $\mathbf{x}_{i,t}$, 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 **?**. 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.

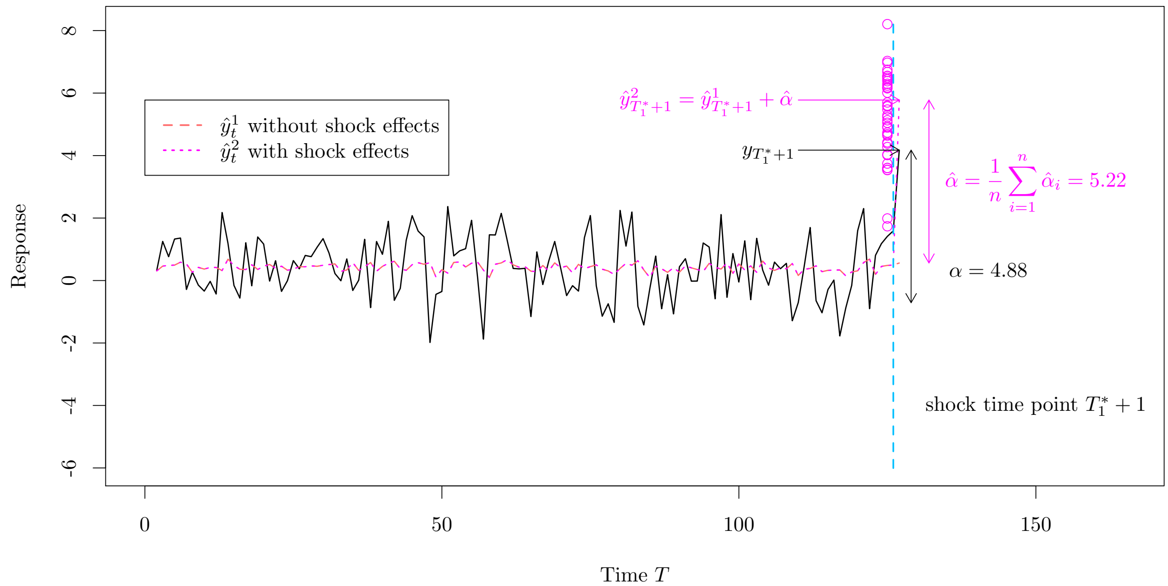

Figure 1. The time series experience a shock at $T_1^* + 1 = 126$ with true shock effect $\alpha = 4.88$. The figure is a comparison between forecast without considering shock effects and the one uses simple averaging given $n = 40$ disparate time series, and that the shock time is at $T_1^* + 1 = 126$. 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 $\hat{y}^1_{T_1^*+1}$.

Figure **??** provides a 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 **??** 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 forecasting 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 structures 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:

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

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

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:

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

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

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

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 the 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_i and δ_i:

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

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:

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

where $\hat{\eta}_1$, $\hat{\phi}_1$, $\hat{\theta}_1$, and $\hat{\beta}_1$ are all ordinary least squares (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 $y_{1,t}$, $t = 1, \ldots T_1^*$. Instead, 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 $\mathcal{M}_{22}$. We specifically focus on predictions for y_{1,T_1^*+1}, the first post-shock response. 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 (see Section **??**). We use these to forecast response values y_{1,T_1^*+1} assuming 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 **??**). 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 that for $\mathcal{M}_{21}$ or $\mathcal{M}_{22}$.

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 (**??**) are OLS estimators of all of the α_is for $i = 2, \ldots, n+1$. 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 **??**.

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

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

where $\hat{\alpha}_i$ 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 $\text{E}(\hat{\alpha}_{\text{IVW}})$ and $\text{Var}(\hat{\alpha}_{\text{IVW}})$ but empirical performance of $\hat{\alpha}_{\text{IVW}}$ is assessed via Monte Carlo simulation (see Section **??**).

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 **?**. 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 **?** to motivate our weighted-adjustment estimator. Consider a $\mathbf{W} \in \mathbb{R}^n$ weight vector $\mathbf{W} = (w_2, \ldots, w_{n+1})'$, where $w_i \in [0, 1]$ for all $i = 2, \ldots, n+1$. Construct

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{x}_{1,T_1^*}' \\ \mathbf{x}_{1,T_1^*+1}' \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_{2,T_2^*}' \\ \mathbf{x}_{2,T_2^*+1}' \\ \vdots \\ \mathbf{x}_{n+1,T_{n+1}^*}' \\ \mathbf{x}_{n+1,T_{n+1}^*+1}' \end{pmatrix}, \quad \text{and} \quad \hat{\mathbf{X}}_1(\mathbf{W}) = (\mathbf{W}' \otimes \mathbf{I}_2)\mathbf{X},$$

where $\otimes$ is the Kronecker product, $\mathbf{I}_2$ is the identity matrix with two rows, and $\mathbf{X}_1, \hat{\mathbf{X}}_1(\mathbf{W}) \in \mathbb{R}^{2\times p}$. Define $\mathcal{W} = \{\mathbf{W} \in [0,1]^n \colon 1_n'\mathbf{W} = 1\}$. 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

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

? commented that we can select $\mathbf{W}^*$ so that (**??**) 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 (**??**), the solution to (**??**) is unique.*

Proposition **??** 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 (**??**) and $2p < n$, there are infinitely many solutions to (**??**). The issue of non-uniqueness is further discussed in Section **??**.

Remark 1. In Section **??** 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 expressions 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 **??**). 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 (**??**). 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 more importantly motivates a bootstrap estimation for evaluation of these conditions. These conditions and bootstrap will be discussed in Section **??** and **??**, respectively. Next, we present the variance expressions for $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$ as below.

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

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

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

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

Formal justification for these results can be found in Appendix. Note that the variances are not comparable in closed-form because of the term $\mathrm{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\}$. This term exists because of the inclusion of the random lagged response in our auto regressive model formulation. Under $\mathcal{M}_{22}$, the expression for $\mathrm{Var}(\alpha_i)$ is not of closed form because γ_i and δ_i may be dependent when they are placed in a random-effects model.

Section **??** details conditions needed for risk-reduction and comparisons of adjustment estimators. These conditions involve variances and expectations which may be difficult to compute in practice. To make use of those properties in practice, estimation is required. Sections **??** and **??** introduce parametric bootstrap and leave-one-out cross validation procedures which prospectively estimate the conditions necessary for risk-reduction without observation of the post-shock response for the time series under study. Our simulations test these procedures.

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

Proposition 3. *Under* $\mathcal{M}_1$*,*

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

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

Proposition **??** 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 **??**, Proposition **??** shows that the risk-reduction condition is

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

Condition (**??**) implies two facts: (1) adjustment (Forecast 2) is preferable to no adjustment (Forecast 1) asymptotically in n whenever $\mu_\alpha \neq 0$ (see Forecast in Section **??**); (2) In finite donor pool settings, adjustment is preferable to no adjustment when μ_α is large relative to its variability and overall regression variability.

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

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

In this case, adjustment is preferable to no adjustment 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 $\mathrm{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\}$.

In fact, we can extend checking risk-reduction conditions to verifying the sign of risk-reduction quantities for generalization. For an adjustment estimator $\hat{\alpha}$, the risk-reduction quantity is defined as $\Delta(\hat{\alpha}) = R_{T_1^*+1,1} - R_{T_1^*+1,2}$. If $\Delta(\hat{\alpha}) > 0$, $\hat{\alpha}$ will improve the Forecast 1. In this context, according to Proposition **??**, under $\mathcal{M}_1$, $\Delta(\hat{\alpha}_{\text{adj}}) = \mu_\alpha^2 - \mathrm{Var}(\hat{\alpha}_{\text{adj}})$ and $\Delta(\hat{\alpha}_{\text{wadj}}) = \mu_\alpha^2 - \mathrm{Var}(\hat{\alpha}_{\text{wadj}})$. Sections **??** and **??** will provide a detailed treatment about how to checking the sign in practice.

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

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

Proposition 4. *If* $\mathbf{W}^*$ *satisfies (**??**), under* $\mathcal{M}_{21}$ *and* $\mathcal{M}_{22}$*,* $R_{T_1^*+1,2} < R_{T_1^*+1,1}$ *when* $\mathrm{Var}(\hat{\alpha}_{\text{wadj}}) < (\mathrm{E}(\alpha_1))^2$.

Under Proposition **??**, we can obtain a risk-reduction inequality that is similar to (**??**),

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

where $\mathrm{Var}(\alpha_i)$ may be replaced with σ_α^2 in $\mathcal{M}_{21}$. The conclusions and intuitions will be identically the same as what we have in Section **??**. Proposition **??** shows that $\hat{\alpha}_{\mathrm{adj}}$ is a biased estimator of $\mathrm{E}(\alpha_1)$ under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$ generally. Hence, Proposition **??** no longer holds for $\hat{\alpha}_{\mathrm{adj}}$ under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$.

As an alternative, we can derive similar risk-reduction conditions that are appropriate for this setting. By Lemma **??** (see Section **??**) and risk decomposition, we will achieve risk-reduction as long as

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

The above inequality simplifies to

$$(\mathrm{E}(\alpha_1))^2 > \mathrm{Var}(\hat{\alpha}_{\mathrm{adj}}) + (\mathrm{E}(\hat{\alpha}_{\mathrm{adj}}) - \mathrm{E}(\alpha_1))^2. \tag{9}$$

As mentioned in Section **??**, it is difficult to evaluate the expectation and variance of $\hat{\alpha}_{\mathrm{IVW}}$. We note that $\hat{\alpha}_{\mathrm{IVW}}$ is generally biased for $\mathrm{E}(\alpha_1)$. That is to say we can adapt the above proof to derive the risk-reduction conditions for $\hat{\alpha}_{\mathrm{IVW}}$: under $\mathcal{M}_1$, $\mathcal{M}_{21}$, and $\mathcal{M}_{22}$, $R_{T_1^*+1,2} < R_{T_1^*+1,1}$ when $\mathrm{Var}(\hat{\alpha}_{\mathrm{IVW}}) + (\mathrm{E}(\hat{\alpha}_{\mathrm{IVW}}) - \mathrm{E}(\alpha_1))^2 < (\mathrm{E}(\alpha_1))^2$. Therefore, using the generalization mentioned in Section **??**, under $\mathcal{M}_2$, it can be shown that the risk-reduction quantities are

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

where we estimate $\Delta(\hat{\alpha})$ using the bootstrap and LOOCV techniques developed in Sections **??** and **??**.

3.2 Bootstrap for risk-reduction evaluation problems

In this section, we present bootstrap procedures that approximate the distribution of our shock-effect estimators, checks the underlying conditions of our risk reduction propositions, and estimate risk-reduction quantity 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 **?**, and it involves the resampling the residuals which are assumed to be the realizations of an iid processe.

Our first bootstrap procedure is as follows: Let B be the bootstrap sample size. At iteration b, first resample the indicies $I = \{2, \ldots, n+1\}$ of the donor pool without replacement to form $I^{(b)}$. Initialize $y_{i,0}$ for all $i \in I^{(b)}$. Then, resample the residuals under models $\mathcal{M}_1$, $\mathcal{M}_{21}$, or $\mathcal{M}_{22}$ and obtain shock-effect estimators for each of the disparate time series for all $i \in I^{(b)}$. These shock effect estimators are then used to construct any of the adjustment estimators $\hat{\alpha}_{\mathrm{adj}}^{(b)}$, $\hat{\alpha}_{\mathrm{wadj}}^{(b)}$, and $\hat{\alpha}_{\mathrm{IVW}}^{(b)}$, for $b = 1, \ldots, B$. We can then estimate distributional quantities of our shock-effect estimators under our considered models with the bootstrap samples $\hat{\alpha}_{\mathrm{adj}}^{(b)}$, $\hat{\alpha}_{\mathrm{wadj}}^{(b)}$, and $\hat{\alpha}_{\mathrm{IVW}}^{(b)}$, for $b = 1, \ldots, B$. We denote this procedure by $\mathcal{B}_u$. We motivate a second bootstrap procedure $\mathcal{B}_f$ which treats the the donor pool as fixed, and not a realization from an infinite super-population. Therefore, there is no resampling of the donor pool in $\mathcal{B}_f$, it is otherwise similar to $\mathcal{B}_u$. An algorithmic formulation of $\mathcal{B}_u$ and $\mathcal{B}_f$ are outlined in Section 2 in the Supplementary Materials.

We will explicitly use these bootstrapped samples of shock-effect estimators to check the risk-reduction conditions in Propositions **??** and **??**. Recall that $\hat{\alpha}_{\mathrm{wadj}}$ and $\hat{\alpha}_{\mathrm{IVW}}$ are unbiased estimators of their expectations, and $\hat{\alpha}_{\mathrm{wadj}}$ is an unbiased estimator of $\mathrm{E}(\alpha_1)$ from Proposition **??**. Our bootstrap procedure estimates the variance of our adjustment estimators. We can then estimate the risk-reduction propositions and inequalities. For example, we can estimate $\Delta(\hat{\alpha}_{\mathrm{adj}})$ under model $\mathcal{M}_{21}$ or $\mathcal{M}_{22}$ with

$$\hat{\Delta}(\hat{\alpha}_{\mathrm{adj}}) = (\hat{\alpha}_{\mathrm{wadj}})^2 - S^2_{\hat{\alpha}_{\mathrm{adj}}} - (\hat{\alpha}_{\mathrm{adj}} - \hat{\alpha}_{\mathrm{wadj}})^2,$$

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

We stress that the our bootstrap approximations cannot alleviate the inherent bias of using our adjustment estimators as surrogates for α_1. 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 $i \in \{2, \ldots, n+1\}$ (**?**); see alternatives for residual bootstrapping in linear models in **?**. Recall that $\mathbf{W}^*$ may not be unique if the conditions in Proposition **??** are not satisfied. Under the setup of our model, non-uniqueness of $\mathbf{W}^*$ would not be a problem for inferential purposes. This is because all risk-reduction propositions and other properties established will still hold. However, non-uniqueness may not be desirable in other model setups. 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 (**??**), say, $\mathbf{W}_1^* = (1, 0)$ and $\mathbf{W}_2^* = (0, 1)$ with $\mathrm{Var}(\alpha_2) \neq \mathrm{Var}(\alpha_3)$. 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 $\hat{\alpha}_{\mathrm{wadj}}$.

3.3 Leave-one-out cross validation

In this section, we adapt leave-one-out cross validation (LOOCV) to our estimation context in order to provide prospective evaluations of our adjustment techniques. Our proposed LOOCV procedure has its roots in Section 7.10 of **?**. Recall in Section **??** that we are given the data $\{(\mathbf{x}_{i,t}, y_{i,t})\colon i = 1, \ldots, n+1, t = 1, \ldots, T_i\}$, where $\{(\mathbf{x}_{1,t}, y_{1,t})\colon t = 1, \ldots, T_1\}$ is the data of the time series of interest and the remaining observations form the donor pool. For iteration $m \in \{1, \ldots, n\}$ of our LOOCV procedure, we set aside $\{(\mathbf{x}_{m+1,t}, y_{m+1,t})\colon t = 1, \ldots, T_{m+1}\}$ as the time series of interest, and construct a new donor pool $\{(\mathbf{x}_{i,t}, y_{i,t})\colon i \in \mathcal{I}_m, t = 1, \ldots, T_i\}$, where $\mathcal{I}_m = \{2, \ldots, n+1\} \setminus \{m+1\}$. Since the post-shock response $y_{m+1, T^*_{m+1}+1}$ is observed, we can evaluate the performance of our adjustment estimators and the original forecast made without adjustment (i.e., Forecast 1 in Section **??**).

LOOCV can be very computationally intensive when n is large, especially when combined with bootstrapping. To alleviate these concerns we can perform LOOCV with a random subset of $k \leq n$ iterations selected without replacement. In this setting, we let $\mathcal{J}$ be the randomly sampled indices. For $m \in \mathcal{J}$, we set aside $\{(\mathbf{x}_{m+1,t}, y_{m+1,t})\colon t = 1, \ldots, T_{m+1}\}$ as the time series of interest, and construct a new donor pool $\{(\mathbf{x}_{i,t}, y_{i,t})\colon i \in \mathcal{I}, t = 1, \ldots, T_i\}$, where $\mathcal{I} = \{2, \ldots, n+1\} \setminus \{m+1\}$. Based on the new donor pool, we estimate relevant parameters using bootstrap procedures outlined in Section **??**. In other words, k times of bootstrapping are nested in a LOOCV procedure. We find that $k = 5$ or $k = 10$ iterations of LOOCV performs well.

We now outline how LOOCV can be used to prospectively assess the performance of adjustment estimators. Let $\mathcal{A}$ be the set of adjustment estimators. For each $\hat{\alpha} \in \mathcal{A}$, let $\delta_{\hat{\alpha}} = I(\hat{\Delta}(\hat{\alpha}) > 0)$ be a decision rule where $I(\cdot)$ is the indicator function and a 1 corresponds to the decision to use adjustment estimator $\hat{\alpha}$. If $\Delta(\hat{\alpha}) > 0$ ($\Delta(\hat{\alpha}) < 0$, respectively) but $\delta_{\hat{\alpha}}$ incorrectly reported 1 (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 incorrect. If $\Delta(\hat{\alpha}) < 0$ ($\Delta(\hat{\alpha}) > 0$, respectively) and $\delta_{\hat{\alpha}}$ correctly reported 0 (1, respectively) so that it make the decision to use $\hat{\alpha}$, $\delta_{\hat{\alpha}}$ is said to be correct. These situations are depicted in the following table:

		Decision	
		$\delta_{\hat{\alpha}} = 1$	$\delta_{\hat{\alpha}} = 0$
Truth	$\Delta(\hat{\alpha}) > 0$	Correct	Incorrect
	$\Delta(\hat{\alpha}) < 0$	Incorrect	Correct

We will use $\mathcal{C}(\delta_{\hat{\alpha}}) = I(\delta_{\hat{\alpha}} \text{ is correct})$ as a metric that evaluates the performance of forecasts made with the adjustment estimator $\hat{\alpha}$. If $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}})) > 0.5$, we claim that $\delta_{\hat{\alpha}}$ is better than random guessing. Note that $\mathcal{C}(\delta_{\hat{\alpha}})$ can generally be computed only when the post-shock response is observed. However, it is possible to estimate $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ using LOOCV. The LOOCV estimates for $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ are

$$\bar{\mathcal{C}}(\delta_{\hat{\alpha}}) = \frac{1}{n}\sum_{m=1}^{n} \mathcal{C}^{(-m)}(\delta_{\hat{\alpha}}), \tag{10}$$

where $\mathcal{C}^{(-m)}(\delta_{\hat{\alpha}})$ is computed with respect to donor pool with index set $\mathcal{I}_m$ and the $m+1$ time series is treated as the time series of interest. The mutual independence assumption in model setup $\mathcal{M}_2$ implies that $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$ will be an almost unbiased estimator of $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ (**?**, Page 222). The LOOCV with k random draws estimates $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ as

$$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}}) = \frac{1}{k}\sum_{m\in\mathcal{J}} \mathcal{C}^{(-m)}(\delta_{\hat{\alpha}}), \tag{11}$$

where $\mathcal{J}$ is the set of the k randomly sampled indicies.

4 Numerical Examples

4.1 Modeling setup

In this section we provide justification for our methods based on Monte Carlo simulation. We implemented our simulation based on $\mathcal{M}_{22}$ with negligibly small Σ_γ and Σ_δ approximating the design of $\mathcal{M}_{21}$. We consider $p = 13$ and $\mu_\alpha = 2$, where $p = 13$ is set to satisfy conditions in Proposition **??**. Parameter setup of our simulations is detailed as follows: the 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, where the minimum allowable value of T_i is fixed to be 90. We will randomly draw T_i^* from $\{2p + 4, \ldots, T_i - 1\}$. The choices of T_i and T_i^* are set up to satisfy a necessary condition for the design matrix of OLS estimation to have full rank. Moreover, it is designed to illustrate the performance of $\hat{\alpha}_{\mathrm{IVW}}$ 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 $\hat{\alpha}_{\mathrm{wadj}}$ may perform well. Last, we set $\gamma_i, \delta_i \overset{iid}{\sim} \mathcal{N}(1, 0.5)$ and $\theta_i, \beta_i \sim \mathcal{N}(0, 1)$. We will 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 choose a Monte Carlo sample size of 30 replications and a bootstrap sample size of $B = 200$ for computation. Means and standard errors for estimated quantities will be recorded. Our LOOCV procedure will consider $k = 5$ random draws. Recall in Section **??** that k times of bootstrap are nested in a LOOCV with k random draws. It implies that $B(k + 1)$ times of bootstrap replications are required for each Monte Carlo simulation.

4.2 Performance metrics

Our adjustment estimators will be evaluated by multiple criteria. We interpret $\delta_{\hat{\alpha}} = I(\hat{\Delta}(\hat{\alpha}) > 0)$ for $\hat{\alpha} \in \mathcal{A}$ as the *guess*, with 1 indicating that $\hat{\alpha}$ provides risk-reduction over the simple no-adjustment forecast, and 0 indicates the converse. We will consider the LOOCV estimators (**??**) and (**??**) to assess correct decision making. We will also consider the Euclidean distance between the post-shock forecasts $\hat{y}_{1,T_1^*+1}$, $\hat{y}_{1,T_1^*+1} + \hat{\alpha}_{\mathrm{adj}}$, $\hat{y}_{1,T_1^*+1} + \hat{\alpha}_{\mathrm{wadj}}$, and $\hat{y}_{1,T_1^*+1} + \hat{\alpha}_{\mathrm{IVW}}$ and the realized post-shock response y_{1,T_1^*+1}.

The first two metrics can combine to assess our forecasting methodology prospectively while the latter requires the realization of the post-shock response y_{1,T_1^*+1}.

4.3 Monte Carlo results

In this section, we discuss simulation results for the bootstrap procedures used in estimating parameters for risk-reduction propositions and inequalities. We mainly discuss simulations under $\mathcal{M}_2$ (see Section **??**) for $\mathcal{B}_u$ and $\mathcal{B}_f$ (see Section **??**) with comparisons to those under $\mathcal{M}_1$ whose results are listed in Section 3 in the Supplementary Materials. Two simulation setups are investigated.

In the first simulation setup, we consider the parameter combination of $n \in \{5, 10, 15, 25\}$ and $\sigma_\alpha \in \{5, 10, 25, 50, 100\}$ where we fix $\sigma = 10$. 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. Results are displayed in Table **??**.

In the second simulation setup, we consider the parameter combination of $\sigma, \sigma_\alpha \in \{5, 10, 25, 50, 100\}$ where we fix $n = 10$. 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. Results are displayed in Table **??**.

Table 1: 30 Monte Carlo simulations of $\mathcal{M}_2$ for $\mathcal{B}_u$ with varying n and σ_α

		Guess			LOOCV with k random draws			Distance to y_{1,T_1^*+1}			
n	σ_α	$\delta_{\hat{\alpha}_{\mathrm{adj}}}$	$\delta_{\hat{\alpha}_{\mathrm{wadj}}}$	$\delta_{\hat{\alpha}_{\mathrm{IVW}}}$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\mathrm{adj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\mathrm{wadj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\mathrm{IVW}}})$	Original	$\hat{\alpha}_{\mathrm{adj}}$	$\hat{\alpha}_{\mathrm{wadj}}$	$\hat{\alpha}_{\mathrm{IVW}}$
	5	1 (0)	1 (0)	1 (0)	0.92 (0.02)	0.96 (0.01)	0.92 (0.02)	65.47 (4.77)	19.92 (3.57)	21.86 (3.99)	20.27 (3.62)
	10	1 (0)	1 (0)	1 (0)	0.9 (0.02)	0.92 (0.02)	0.9 (0.02)	65.67 (4.9)	20.66 (3.74)	22.92 (4.2)	21.1 (3.75)
5	25	0.97 (0.03)	1 (0)	0.97 (0.03)	0.8 (0.02)	0.81 (0.02)	0.8 (0.03)	66.29 (5.83)	27.98 (4.21)	30.3 (4.87)	28.3 (4.18)
	50	0.83 (0.07)	0.87 (0.06)	0.87 (0.06)	0.55 (0.05)	0.57 (0.04)	0.55 (0.05)	70.59 (7.46)	43.1 (6.19)	46.31 (6.8)	43.07 (6.16)
	100	0.47 (0.09)	0.73 (0.08)	0.47 (0.09)	0.48 (0.04)	0.48 (0.04)	0.46 (0.04)	89.82 (10.49)	76.74 (11.42)	79.78 (12.26)	76.69 (11.33)
	5	1 (0)	1 (0)	1 (0)	0.95 (0.02)	0.95 (0.02)	0.95 (0.02)	55.66 (4.28)	16.36 (2.48)	17.51 (2.41)	16.6 (2.45)
	10	1 (0)	1 (0)	1 (0)	0.92 (0.02)	0.91 (0.03)	0.92 (0.02)	55.91 (4.71)	18.28 (2.85)	19.38 (2.88)	18.44 (2.83)
10	25	0.9 (0.06)	0.97 (0.03)	0.93 (0.05)	0.77 (0.04)	0.79 (0.04)	0.75 (0.04)	59.43 (6.1)	29.01 (4.24)	32.1 (4.3)	28.84 (4.26)
	50	0.77 (0.08)	0.8 (0.07)	0.77 (0.08)	0.55 (0.04)	0.64 (0.04)	0.55 (0.04)	69.34 (9.45)	52.52 (6.75)	58.05 (7.28)	52.28 (6.76)
	100	0.63 (0.09)	0.7 (0.09)	0.63 (0.09)	0.53 (0.04)	0.53 (0.05)	0.51 (0.04)	104.19 (15.88)	99.93 (12.83)	113.77 (13.42)	99.57 (12.81)
	5	1 (0)	1 (0)	1 (0)	0.92 (0.02)	0.93 (0.02)	0.92 (0.02)	51.78 (2.74)	12.66 (2.48)	13.89 (2.48)	12.64 (2.49)
	10	1 (0)	1 (0)	1 (0)	0.92 (0.02)	0.91 (0.02)	0.92 (0.02)	51.82 (3.04)	13.93 (2.52)	16 (2.63)	14.05 (2.53)
15	25	1 (0)	1 (0)	1 (0)	0.83 (0.03)	0.83 (0.03)	0.83 (0.03)	51.94 (4.56)	21.27 (2.93)	24.49 (3.69)	21.75 (2.92)
	50	0.9 (0.06)	0.93 (0.05)	0.9 (0.06)	0.71 (0.03)	0.67 (0.04)	0.69 (0.04)	55.22 (7.06)	35.85 (4.63)	42.72 (5.93)	36.41 (4.71)
	100	0.7 (0.09)	0.73 (0.08)	0.7 (0.09)	0.54 (0.05)	0.61 (0.04)	0.54 (0.05)	76.44 (10.75)	67.79 (8.77)	79.49 (11.44)	68.57 (8.96)
	5	1 (0)	1 (0)	1 (0)	0.9 (0.03)	0.91 (0.02)	0.9 (0.03)	62.23 (6.88)	21.29 (5.9)	19.29 (5.88)	21.34 (5.93)
	10	1 (0)	1 (0)	1 (0)	0.87 (0.03)	0.89 (0.02)	0.87 (0.03)	61.28 (7.09)	22.96 (5.91)	20.81 (5.89)	23.02 (5.94)
25	25	1 (0)	1 (0)	1 (0)	0.73 (0.03)	0.74 (0.04)	0.73 (0.03)	61.9 (7.5)	30.04 (6.46)	28.76 (6.46)	30.17 (6.48)
	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)	67.46 (9.23)	45.5 (8.55)	47.38 (8.62)	45.67 (8.58)
	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)	91.08 (14.54)	81.72 (14.24)	85.72 (15.39)	82.3 (14.24)

First, assuming that $\bar{C}^{(k)}(\delta_{\hat{\alpha}})$ well estimates $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ and fixing n, we observe from Table **??** that the decision making of $\delta_{\hat{\alpha}}$ is nearly correct for $\hat{\alpha} \in \mathcal{A}$ when σ_α is small from Table **??**. The reasons can be explained as follows. When σ_α is small, the signal of the covariates is strong so that $\hat{\alpha}_{\mathrm{wadj}}$ will be expected to capture the signal according to construction of $\hat{\alpha}_{\mathrm{wadj}}$ in Section **??**. 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 **??**. However, when the signal of the covariates is poor (σ_α is big), the decision rule $\delta_{\hat{\alpha}}$ becomes unreliable 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 the correctness of $\delta_{\hat{\alpha}}$ increases when n increases. It is due to the robustness gain in estimation when n increases.

Additionally, we observe that in most cases $\delta_{\hat{\alpha}_{\mathrm{wadj}}}$ reports $\hat{\alpha}_{\mathrm{wadj}}$ reduces the risk even when $\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\mathrm{wadj}}})$ starts to break down. Recall from Section **??** that $\hat{\Delta}(\hat{\alpha})$ contains the squared bias for estimating $\mathrm{E}(\alpha_1)$. But it is not present for $\hat{\Delta}(\hat{\alpha}_{\mathrm{wadj}})$ since we applied the fact $\hat{\alpha}_{\mathrm{wadj}}$ is unbiased for $\mathrm{E}(\alpha_1)$ from Proposition **??** in plugging it in with replacing $\mathrm{E}(\alpha_1)$. Therefore, when the signal from covariates is poorer, $\delta_{\hat{\alpha}_{\mathrm{wadj}}}$ 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 correctness 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 $\hat{\alpha}_{\mathrm{adj}}$ reduces the risk is approximately $0.83 \times 0.55 = 0.457$ from Table **??**. 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 $\hat{\alpha} \in \mathcal{A}$.

From columns related to distance to y_{1,T_1^*+1} in Table **??**, as σ_α increases, the prediction appears to be poorer. When $\sigma_\alpha = 5, 10, 25$, forecasts using $\hat{\alpha}_{\mathrm{adj}}$, $\hat{\alpha}_{\mathrm{wadj}}$, and $\hat{\alpha}_{\mathrm{IVW}}$ 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, no statistical evidence has been found to support the claim that n matters in prediction. In other words, the size of the donor pool matters for producing reliable decision-making of $\delta_{\hat{\alpha}}$ rather than reliable prediction.

Table 2: 30 Monte Carlo simulations of $\mathcal{M}_2$ for $\mathcal{B}_u$ with varying σ and σ_α

		Guess			LOOCV with k random draws			Distance to y_{1,T_1^*+1}			
σ	σ_α	$\delta_{\hat{\alpha}_{\mathrm{adj}}}$	$\delta_{\hat{\alpha}_{\mathrm{wadj}}}$	$\delta_{\hat{\alpha}_{\mathrm{IVW}}}$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\mathrm{adj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\mathrm{wadj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\mathrm{IVW}}})$	Original	$\hat{\alpha}_{\mathrm{adj}}$	$\hat{\alpha}_{\mathrm{wadj}}$	$\hat{\alpha}_{\mathrm{IVW}}$
	5	1 (0)	1 (0)	1 (0)	0.97 (0.01)	0.97 (0.01)	0.97 (0.01)	49.95 (2.46)	12.36 (1.71)	11.71 (1.55)	12.13 (1.71)
	10	1 (0)	1 (0)	1 (0)	0.92 (0.02)	0.92 (0.02)	0.92 (0.02)	51.25 (2.88)	13.14 (1.88)	14.55 (1.7)	12.84 (1.87)
5	25	0.97 (0.03)	1 (0)	1 (0)	0.81 (0.03)	0.84 (0.03)	0.81 (0.03)	54.85 (4.6)	19.32 (3.07)	20.87 (3.45)	19.75 (3.13)
	50	0.77 (0.08)	0.8 (0.07)	0.77 (0.08)	0.59 (0.04)	0.62 (0.03)	0.58 (0.04)	60.57 (8.15)	43.17 (5.49)	47.62 (5.88)	43.59 (5.63)
	100	0.67 (0.09)	0.63 (0.09)	0.67 (0.09)	0.45 (0.04)	0.49 (0.04)	0.49 (0.04)	71.15 (10.52)	72.15 (9.61)	80.17 (9.52)	71.7 (9.57)
	5	0.97 (0.03)	1 (0)	0.97 (0.03)	0.91 (0.02)	0.91 (0.02)	0.91 (0.02)	48.65 (3.85)	18.3 (2.4)	19.78 (2.31)	18.64 (2.39)
	10	1 (0)	1 (0)	1 (0)	0.9 (0.02)	0.93 (0.02)	0.91 (0.02)	52.46 (4.4)	16.52 (3.06)	15.86 (3.18)	16.46 (3.06)
10	25	1 (0)	1 (0)	1 (0)	0.79 (0.04)	0.81 (0.03)	0.8 (0.04)	64.81 (4.93)	24.85 (3.5)	24.78 (3.7)	24.94 (3.58)
	50	0.87 (0.06)	0.9 (0.06)	0.87 (0.06)	0.57 (0.04)	0.61 (0.04)	0.57 (0.04)	65.69 (6.74)	40.16 (5.47)	37.29 (5.35)	39.44 (5.49)
	100	0.63 (0.09)	0.67 (0.09)	0.63 (0.09)	0.43 (0.04)	0.47 (0.03)	0.43 (0.04)	67.98 (10.7)	65.69 (8.35)	78.45 (9.42)	67.75 (8.58)
	5	1 (0)	1 (0)	1 (0)	0.73 (0.03)	0.74 (0.03)	0.73 (0.03)	58.32 (7.69)	39.37 (7.46)	39.16 (7.89)	39.6 (7.47)
	10	1 (0)	1 (0)	1 (0)	0.71 (0.04)	0.73 (0.05)	0.73 (0.04)	65.93 (5.72)	33.34 (5.91)	34.04 (5.49)	33.05 (5.89)
25	25	0.9 (0.06)	0.97 (0.03)	0.93 (0.05)	0.67 (0.04)	0.68 (0.03)	0.66 (0.04)	56.05 (7.5)	39.49 (5.25)	35.65 (4.98)	39.73 (5.26)
	50	0.77 (0.08)	0.8 (0.07)	0.77 (0.08)	0.59 (0.04)	0.62 (0.04)	0.59 (0.04)	60.85 (7.93)	46.55 (6.83)	47.97 (7.4)	47.23 (6.73)
	100	0.67 (0.09)	0.7 (0.09)	0.63 (0.09)	0.53 (0.05)	0.53 (0.04)	0.53 (0.05)	95.21 (12.76)	98.68 (11.57)	99.87 (13.28)	98.77 (11.61)
	5	0.77 (0.08)	0.7 (0.09)	0.77 (0.08)	0.6 (0.04)	0.61 (0.03)	0.61 (0.04)	71.05 (8.49)	52.9 (8.7)	57.45 (9.57)	53.05 (8.59)
	10	0.63 (0.09)	0.63 (0.09)	0.63 (0.09)	0.57 (0.04)	0.57 (0.05)	0.57 (0.04)	68.23 (7.38)	44.22 (5.95)	55.53 (6.2)	44.63 (5.86)
50	25	0.73 (0.08)	0.67 (0.09)	0.77 (0.08)	0.55 (0.04)	0.55 (0.04)	0.57 (0.04)	69.78 (10.51)	63.97 (9.49)	68.82 (9.79)	64.33 (9.47)
	50	0.83 (0.07)	0.8 (0.07)	0.83 (0.07)	0.55 (0.04)	0.51 (0.05)	0.55 (0.04)	64.9 (11.37)	66.61 (10.07)	73.22 (10.92)	67.06 (9.99)
	100	0.47 (0.09)	0.53 (0.09)	0.47 (0.09)	0.49 (0.04)	0.51 (0.04)	0.49 (0.04)	92.78 (12)	73.61 (11.81)	78.33 (10.92)	74.37 (11.63)
	5	0.5 (0.09)	0.47 (0.09)	0.47 (0.09)	0.49 (0.05)	0.48 (0.04)	0.47 (0.05)	125.13 (15.27)	104.01 (14.95)	100.33 (13.71)	105.4 (14.95)
	10	0.4 (0.09)	0.4 (0.09)	0.37 (0.09)	0.49 (0.03)	0.51 (0.05)	0.51 (0.04)	106.46 (14.76)	101.82 (14.64)	110.6 (14.95)	98.85 (14.69)
100	25	0.63 (0.09)	0.57 (0.09)	0.63 (0.09)	0.49 (0.04)	0.48 (0.04)	0.51 (0.04)	142.91 (16.43)	132.29 (18.34)	146.26 (17.96)	132.06 (18.25)
	50	0.57 (0.09)	0.57 (0.09)	0.57 (0.09)	0.5 (0.05)	0.5 (0.05)	0.49 (0.04)	114.36 (15.46)	93 (16.26)	95.41 (15.89)	91.92 (15.98)
	100	0.33 (0.09)	0.3 (0.09)	0.3 (0.09)	0.49 (0.05)	0.49 (0.05)	0.48 (0.05)	150.5 (20.71)	148.07 (17.57)	151.26 (17.18)	148.96 (17.71)

From Table **??**, we observe that as σ_α increases fixing σ, $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$ decreases, which is a pattern similar to the one shown in the first experiment. Furthermore, as σ increases fixing σ_α, $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$ decreases as well. Note that the correctness hinges on the estimation of the parameters. Since $\hat{\alpha}_{\mathrm{wadj}}$ is a linear combination of OLS estimates, as σ increases, $\mathrm{Var}(\hat{\alpha}_{\mathrm{wadj}})$ increases as well. Therefore, $\hat{\alpha}_{\mathrm{wadj}}$ become more volatile and its estimation of $\mathrm{E}(\alpha_1)$ can be less reliable. Those reasons can explain why an increase of σ_α contributes to a decrease of $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$. We observe similar patterns for distance to y_{1,T_1^*+1} as well. 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 $\hat{\alpha} \in \mathcal{A}$.

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

Table 3: 30 Monte Carlo simulations of $\mathcal{M}_2$ for $\mathcal{B}_f$ with varying n and σ_α

		Guess			LOOCV with k random draws			Distance to y_{1,T_1^*+1}			
n	σ_α	$\delta_{\hat{\alpha}_{\text{adj}}}$	$\delta_{\hat{\alpha}_{\text{wadj}}}$	$\delta_{\hat{\alpha}_{\text{IVW}}}$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{adj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{wadj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{IVW}}})$	Original	$\hat{\alpha}_{\text{adj}}$	$\hat{\alpha}_{\text{wadj}}$	$\hat{\alpha}_{\text{IVW}}$
	5	1 (0)	1 (0)	1 (0)	0.89 (0.02)	0.9 (0.02)	0.89 (0.02)	51.92 (4.04)	19.23 (2.55)	20.64 (2.76)	19.36 (2.52)
	10	1 (0)	1 (0)	1 (0)	0.89 (0.02)	0.89 (0.02)	0.89 (0.02)	52.58 (4.35)	21.57 (2.58)	23.38 (2.86)	21.72 (2.5)
5	25	1 (0)	1 (0)	1 (0)	0.78 (0.03)	0.83 (0.02)	0.79 (0.03)	55.01 (5.84)	30.3 (3.59)	31.98 (4.36)	30.2 (3.5)
	50	0.9 (0.06)	0.97 (0.03)	0.93 (0.05)	0.66 (0.04)	0.65 (0.04)	0.65 (0.04)	64.42 (8.11)	50 (5.78)	52.56 (6.92)	49.55 (5.7)
	100	0.77 (0.08)	0.97 (0.03)	0.8 (0.07)	0.57 (0.04)	0.53 (0.04)	0.57 (0.04)	91.51 (13.2)	94.26 (10.41)	97.26 (12.57)	93.61 (10.27)
	5	1 (0)	1 (0)	1 (0)	0.91 (0.02)	0.92 (0.02)	0.91 (0.02)	52.34 (4)	17.23 (2.96)	18.54 (2.83)	17.39 (2.95)
	10	1 (0)	1 (0)	1 (0)	0.89 (0.02)	0.89 (0.02)	0.89 (0.02)	52.59 (4.04)	19.02 (3.24)	20.95 (3.13)	19.23 (3.24)
10	25	1 (0)	1 (0)	1 (0)	0.75 (0.03)	0.78 (0.03)	0.77 (0.03)	54.07 (4.97)	27.66 (4.55)	31.79 (4.47)	27.84 (4.61)
	50	0.83 (0.07)	1 (0)	0.8 (0.07)	0.59 (0.04)	0.63 (0.04)	0.59 (0.04)	60.32 (7.53)	47.7 (7.04)	52.97 (7.54)	47.78 (7.17)
	100	0.8 (0.07)	0.93 (0.05)	0.8 (0.07)	0.47 (0.04)	0.51 (0.04)	0.46 (0.04)	85.6 (12.99)	89.85 (12.82)	100.74 (13.61)	90.4 (12.91)
	5	1 (0)	1 (0)	1 (0)	0.91 (0.02)	0.93 (0.02)	0.91 (0.02)	49.85 (4.01)	18.07 (2.88)	18.38 (2.71)	18.04 (2.88)
	10	1 (0)	1 (0)	1 (0)	0.87 (0.02)	0.89 (0.02)	0.87 (0.02)	48.73 (4.3)	19.45 (2.97)	19.35 (2.86)	19.32 (2.99)
15	25	1 (0)	1 (0)	1 (0)	0.75 (0.03)	0.78 (0.03)	0.76 (0.03)	47.06 (5.13)	26.16 (3.31)	26.81 (3.13)	26.23 (3.33)
	50	0.93 (0.05)	1 (0)	0.9 (0.06)	0.61 (0.04)	0.69 (0.03)	0.64 (0.04)	48.75 (6.86)	40.27 (4.43)	42.09 (4.49)	40.77 (4.38)
	100	0.67 (0.09)	1 (0)	0.63 (0.09)	0.55 (0.04)	0.51 (0.04)	0.55 (0.04)	64.29 (11.11)	68.85 (8.27)	74.08 (8.72)	69.91 (8.14)
	5	1 (0)	1 (0)	1 (0)	0.95 (0.02)	0.94 (0.02)	0.95 (0.02)	57.6 (6.94)	21.58 (5.9)	20 (5.86)	21.58 (5.9)
	10	1 (0)	1 (0)	1 (0)	0.93 (0.02)	0.91 (0.02)	0.93 (0.02)	56.8 (7.01)	22.2 (5.97)	20.47 (5.89)	22.22 (5.97)
25	25	1 (0)	1 (0)	1 (0)	0.78 (0.04)	0.79 (0.04)	0.77 (0.04)	56.58 (7.49)	28.96 (6.49)	27.06 (6.19)	29.03 (6.48)
	50	0.9 (0.06)	1 (0)	0.9 (0.06)	0.57 (0.04)	0.6 (0.04)	0.58 (0.04)	64.33 (8.75)	47.16 (8.28)	46.3 (7.26)	47.35 (8.25)
	100	0.83 (0.07)	1 (0)	0.8 (0.07)	0.49 (0.04)	0.48 (0.04)	0.5 (0.04)	95.61 (13.02)	90.1 (13.29)	86.81 (11.75)	90.55 (13.23)

Table 4: 30 Monte Carlo simulations of $\mathcal{M}_2$ for $\mathcal{B}_f$ with varying σ and σ_α

		Guess			LOOCV with k random draws			Distance to y_{1,T_1^*+1}			
σ	σ_α	$\delta_{\hat{\alpha}_{\text{adj}}}$	$\delta_{\hat{\alpha}_{\text{wadj}}}$	$\delta_{\hat{\alpha}_{\text{IVW}}}$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{adj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{wadj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{IVW}}})$	Original	$\hat{\alpha}_{\text{adj}}$	$\hat{\alpha}_{\text{wadj}}$	$\hat{\alpha}_{\text{IVW}}$
	5	1 (0)	1 (0)	1 (0)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	53.27 (2.59)	10.91 (1.7)	11.77 (1.55)	10.8 (1.69)
	10	1 (0)	1 (0)	1 (0)	0.91 (0.02)	0.95 (0.02)	0.92 (0.02)	50.86 (3.83)	18.06 (2.18)	17.26 (2.17)	18.25 (2.17)
5	25	0.97 (0.03)	1 (0)	0.97 (0.03)	0.83 (0.03)	0.89 (0.02)	0.83 (0.03)	60.52 (3.97)	16.09 (2.4)	19.83 (3.05)	15.87 (2.38)
	50	0.87 (0.06)	1 (0)	0.87 (0.06)	0.67 (0.04)	0.69 (0.03)	0.67 (0.04)	54.65 (6.95)	51.79 (6.87)	53.83 (7.26)	52.6 (6.91)
	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)	104.73 (13)	88.31 (12.16)	88.88 (12.63)	86.72 (12.35)
	5	1 (0)	1 (0)	1 (0)	0.93 (0.02)	0.93 (0.02)	0.93 (0.02)	58.17 (4.18)	18.17 (2.61)	16.59 (2.39)	18.09 (2.63)
	10	1 (0)	1 (0)	1 (0)	0.86 (0.03)	0.89 (0.03)	0.87 (0.03)	52.81 (4.07)	19.05 (2.49)	19.67 (2.66)	18.95 (2.53)
10	25	0.97 (0.03)	1 (0)	0.97 (0.03)	0.78 (0.03)	0.82 (0.03)	0.79 (0.03)	61.53 (5.69)	28.55 (4.07)	31.82 (4.27)	28.76 (4.07)
	50	0.93 (0.05)	1 (0)	0.9 (0.06)	0.66 (0.04)	0.67 (0.04)	0.66 (0.04)	56.31 (8.03)	47.33 (4.89)	41.44 (4.39)	46.86 (4.79)
	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)	84.38 (11.62)	82.91 (11.77)	84.5 (12.39)	84.3 (11.93)
	5	0.97 (0.03)	1 (0)	1 (0)	0.78 (0.03)	0.77 (0.03)	0.79 (0.03)	56.35 (6.1)	25.54 (4.06)	30.49 (4.61)	25.94 (4.02)
	10	1 (0)	1 (0)	1 (0)	0.81 (0.03)	0.79 (0.03)	0.81 (0.03)	49.8 (5.24)	25.04 (4.38)	26.74 (3.64)	24.86 (4.38)
25	25	1 (0)	1 (0)	1 (0)	0.72 (0.03)	0.77 (0.03)	0.71 (0.03)	54.21 (6.51)	44.17 (6.66)	43.41 (6.95)	43.89 (6.63)
	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)	66.51 (7.86)	46.19 (8.04)	44.83 (9.31)	46.54 (8.27)
	100	0.9 (0.06)	1 (0)	0.87 (0.06)	0.54 (0.04)	0.54 (0.04)	0.56 (0.04)	109.21 (13.29)	78.47 (12.45)	83.99 (12)	78.57 (12.75)
	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)	75.29 (10.75)	63.51 (8.3)	64.25 (8.88)	63.32 (8.4)
	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)	57.59 (6)	48.51 (8.32)	50.58 (8.63)	48.08 (8.36)
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)	77.21 (12.03)	54.73 (9.47)	54.76 (10.24)	54.2 (9.62)
	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)	90.48 (10.21)	68.88 (9.2)	68.28 (10.48)	68.8 (9.28)
	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)	111.09 (17.54)	102.47 (16.68)	110.15 (15.87)	101.53 (16.52)
	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)	214.07 (67.4)	195.46 (67.65)	197.51 (68.45)	196.58 (67.51)
	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)	120.85 (15.39)	114.79 (15.51)	119.19 (15.75)	114.02 (15.74)
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)	97.84 (13.75)	95.02 (14.88)	100.04 (16.94)	96.23 (14.71)
	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)	141.22 (24.51)	136.11 (25.73)	150.62 (27.62)	135.26 (25.39)
	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)	95.53 (12.8)	103.01 (14.12)	103.31 (16.27)	105.79 (14.32)

Simulation for $\mathcal{B}_f$ with the same parameter setup as that of $\mathcal{B}_u$ are implemented. See Table **??** and Table **??** for results. Comparing Table **??** and Table **??** 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 $\mathcal{B}_f$ with statistical evidence. For other situations, $\mathcal{B}_u$ and $\mathcal{B}_f$ 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 **??** and Table **??**, it appears that when $n = 10$ and $\sigma_\alpha = 5$, $\mathcal{B}_u$ is better than $\mathcal{B}_f$ 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 $n = 10$, $\mathcal{B}_f$ and $\mathcal{B}_u$ are quite similar under situations of different σ and σ_α. In conclusion, $\mathcal{B}_u$ is better than $\mathcal{B}_f$ when the signal of the covariates is strong and n is small; otherwise, they are similar.

Simulation results corresponding to model $\mathcal{M}_1$ are listed in Section 3 in Supplementary Materials. Results under model $\mathcal{M}_1$ are very similar to those of $\mathcal{M}_2$, except for the difference among estimators. The results show that (1) the performance of $\hat{\alpha}_{\mathrm{adj}}$ and $\hat{\alpha}_{\mathrm{IVW}}$ are nearly the same and (2) in many situations, $\hat{\alpha}_{\mathrm{adj}}$ and $\hat{\alpha}_{\mathrm{IVW}}$ are better than $\hat{\alpha}_{\mathrm{wadj}}$; in other situations, they are mostly the same. Recall that in $\mathcal{M}_1$, the models for α_1 do not involve the covariates. Therefore, similarity weighting may not be informative when the model for α_i is identified wrongly. Under $\mathcal{M}_1$, simple averaging, aimed for a reduction of variance, or inverse-variance weighting, targeting on reducing negative effects from varying time lengths, may work better.

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 (**?**). Focus on the oil sector is because oil prices have been shown to exhibit a cointegrating behavior with economic indices (**?**), 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 on the Sunday before trading resumes on Monday, March 9th (**?**). Furthermore, fear 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 (**?**). New York declared a state of emergency on March 7th (**?**), and by Sunday, March 8th, eight states have declared a state of emergency (**?**) while Italy placed 16 million people in quarantine (**?**).

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 (**?**). An inversion of the yield curve is an event that signals likely recessions (**??**). 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 (**?**). 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 (**?**). 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 (**?**). 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 **??** 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 consists 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 (**?**), and the third was an OPEC induced supply side shock effect (**?**). 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 (**?**).

(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 (**?**). 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 (**??**).

(c) On November 27th, 2014, it is documented that oil prices fall as OPEC opts not to cut production (**?**). 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 (**?**, 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 the Section 1 in the Supplementary Materials. Under $\mathcal{M}_2$, we computed $\hat{\alpha}_{\text{adj}}$, weighted adjustment $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$. For $\hat{\alpha}_{\text{wadj}}$, we observe that $\mathbf{W}^* = (0.000, 0.000, 0.000, 0.000, 1.000)$ and $\left\| \text{vec}\left(\mathbf{X}_1 - \hat{\mathbf{X}}_1(\mathbf{W}^*)\right) \right\|_4 = 1314.04$. Note that the norm is computed using the k-dimensional Euclidean metric. The solution $\mathbf{W}^*$ suggests that the shock effect of interest is replicated by the November 27, 2014 shock effect.

Using the bootstrap procedure $\mathcal{B}_f$, we estimated parameters for risk-reduction propositions and risk-reduction quantities proposed in Section **??**. The estimated bootstrap variances for $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ are 0.531, 0.479, and 0.970, respectively. Plugging these estimates into conditions in Section **??** 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. We verify the consistency of the result yielded by risk-reduction propositions with the reality as below.

We can see from Figure **??** that $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$ and $\hat{\alpha}_{\text{IVW}}$ perform decently well, and 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, **?** 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

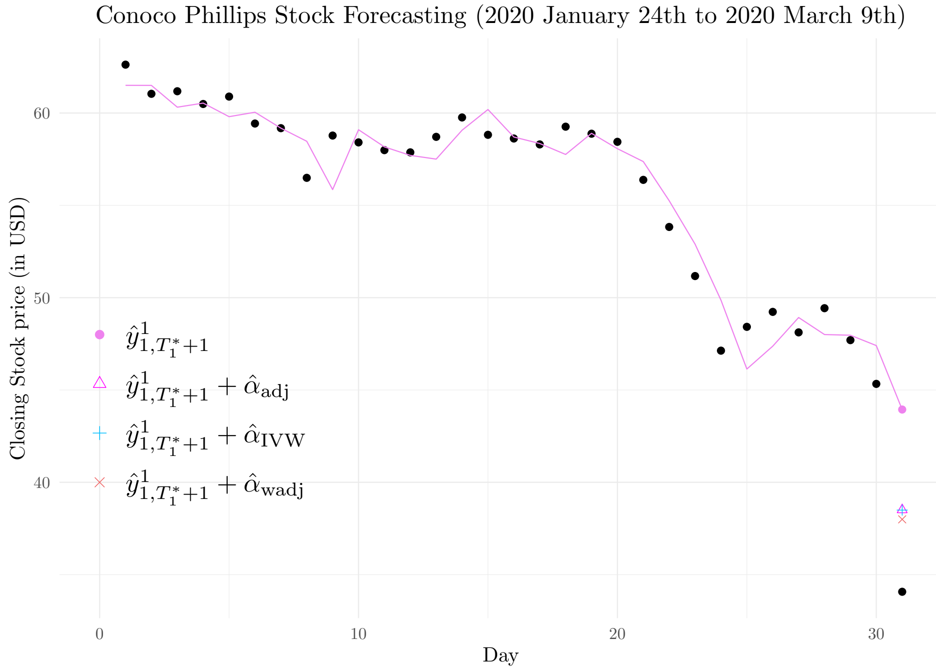

Figure 2: March 9th, 2020 post-shock forecasts for Conoco Phillips stock price.

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.

6 Discussion

We developed methodology for forecasting post-shock response values after the occurrence of a structural shock. Our methodology is as follows: construct a synthetic panel of disparate time series which have undergone similar shocks, estimate the shock-effects in those series, aggregate them, and then adjust the original forecast by adding the aggregated shock-effect estimator to the original forecast. We provided risk-reduction propositions and empirical tools that can prospectively assess the effectiveness of our adjustment strategies in additive shock-effect settings. The model, under which we verify these claims, is a simple AR(1) model. Similar results can be obtained for more general models such as AR(p), vector autoregression, and generalized autoregressive conditional heteroskedasticity models.

Moreover, the mean function for α_i under $\mathcal{M}_2$ can be extended to include more lagged predictors. The functional form of the mean function in our considered models can be extended beyond the linear regression model. Multiple shock-effects can be nested within a time series; and time series in the donor pool can 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 provided that its first and second moments exist. 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 **?** motivated a stationary bootstrap method for strictly stationary and weakly dependent time series $\{X_n : n \in \mathbb{N}\}$. 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 $\{I_n : n \in \mathbb{N}\}$ is a sequence of i.i.d. discrete uniform random variables; and the collection of random lengths $\{L_n : n \in \mathbb{N}\}$ 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 (**?**, 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 (**??**). More work related to bootstrapping time series can be referred to Chapters 3 and 4 in **?**, and **?**. It is up to users in terms of selecting which procedure to choose but under different assumptions on the time series.

We have implicitly assumed that $\mathbf{W}^*$ is non-degenerate in the population. Recall that in Section **??** we noted that if there exists some $\mathbf{W}^*$ which satisfies (**??**) and $2p < n$, then there will be infinitely many solutions to $\mathbf{W}^*$. However, in applications, it is possible that $\mathbf{W}^*$ may take values on the boundary of $\mathcal{W}$, in which case bootstrapping may fail to estimate the distribution of $\hat{\alpha}_{\text{wadj}}$ (**?**). Moreover, when $2p < n$, $\mathcal{B}_u$ fails since the existence of infinitely many solutions is certain (if there exists some $\mathbf{W}^*$ satisfies (**??**)), and will guarantee degeneracy of $\mathbf{W}^*$. However, this issue will not occur under $\mathcal{B}_f$ since it takes $\mathbf{W}^*$ as being fixed and the parameter space is Θ that does not involve the constrained $\mathcal{W}$. Nevertheless, it does not seriously compromise the inference according to our simulation results in Section 4 in the Supplementary Materials. Note that there are some philosophical distinctions between $\mathcal{B}_u$ and $\mathcal{B}_f$. $\mathcal{B}_u$ treats the donor pool as realizations from some infinite super-population of potential donors. In contrast, $\mathcal{B}_f$ treats the donor pool as being fixed and known before the analysis is conducted, where the randomness comes from parameters and idiosyncratic error.

A double bootstrap procedure with similar steps to the bootstrap technique in Section **??** can estimate the distribution of $\hat{\Delta}(\hat{\alpha})$ for $\hat{\alpha} \in \mathcal{A}$. The double bootstrap, instead of checking whether $\Delta(\hat{\alpha}) > 0$, can check whether a bootstrap percentile interval of resampled estimates of $\Delta(\hat{\alpha})$ contain 0 at a desired error threshold. We investigated such a double bootstrap procedure and found that it produced inferences that were similar to those produced using the bootstrap techniques developed in the main text.

There have been several recent time series pooling methods developed for forecasting COVID-19 cases. **?** 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 appropriate for post-shock forecasting but may be too dependent upon model specification for the shock distribution. **?** 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 a clustering approach. 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. **?** 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. Although the methodology in **?** is quite general, there is no guarantee for theoretical properties in prediction without assuming any distributional structure.

7 Appendix

7.1 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 **??**),

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

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

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

Similarly,

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

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

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

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

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

$$\begin{aligned}\mathrm{Var}(\hat{\alpha}_{\text{wadj}}) &= \mathrm{E}\left\{\sum_{i=2}^{n+1}(w_i^*)^2(\sigma^2(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22})\right\} + \sigma_\alpha^2\sum_{i=2}^{n+1}(w_i^*)^2\\ &= \sigma^2\sum_{i=2}^{n+1}(w_i^*)^2\mathrm{E}\big\{(\mathbf{U}_i'\mathbf{U}_i)^{-1}_{22}\big\} + \sigma_\alpha^2\sum_{i=2}^{n+1}(w_i^*)^2.\end{aligned} \qquad (12)$$

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

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

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

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

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

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

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

7.3 Proofs for lemmas and propositions

Proof of Proposition ?? The proof of **?** in Appendix A.2 and A.3 adapts easily to Proposition **??**. $\square$

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

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

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

Lemma 1. *The forecast risk difference is* $R_{T_1^*+1,1} - R_{T_1^*+1,2} = \mathrm{E}(\alpha_1^2) - \mathrm{E}(\hat{\alpha} - \alpha_1)^2$ *for all estimators of* α_1 *that are independent of* Θ_1 *(see Section **??**).*

Proof of Lemma ?? 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 (**??**). Notice that

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

It follows that

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

Assuming $\mathbf{S} = (\mathbf{1}_n, \mathbf{y}_{1,t-1}, \mathbf{x}_1, \mathbf{x}_{1,t-1})$ has full rank, under OLS setting, $\hat{\eta}_1$, $\hat{\phi}_1$, $\hat{\theta}_1$, and $\hat{\beta}_1$ are unbiased estimators of η_1, ϕ_1, θ_1, and β_1, respectively under conditioning of Θ_1. Since we assume $\hat{\alpha}$ is independent of Θ_1, through the method of iterated expectation,

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

It follows that

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

which finishes the proof. $\square$

Proof of Proposition ?? 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 **??**,

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

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 **??**, 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 ?? By Proposition **??**, 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 **??**. □