Minimizing post-shock forecasting error through aggregation of outside information

Abstract

We develop a forecasting methodology for providing credible forecasts for time series that have recently undergone a shock. We achieve this by borrowing knowledge from other time series that have undergone similar shocks for which post-shock outcomes are observed. Three shock effect estimators are motivated with the aim of minimizing average forecast risk. We propose risk-reduction propositions that provide conditions that establish when our methodology works. Bootstrap and leave-one-out cross validation procedures are provided to prospectively assess the performance of our methodology. Several simulated data examples, and two real data examples of forecasting Conoco Phillips and Apple stock price are provided for verification and illustration.

Keywords: Data Integration, Prospective forecasting, Risk reduction, Residual bootstrap, Cross validation

1. Introduction

We provide forecasting adjustment techniques with the goal of lowering overall forecast error when the time series under study has undergone a structural shock. We focus on the specific setting in which a structural shock has occurred and one desires a prediction for the post-shock response at the next time point. Standard forecasting methods may not yield accurate predictions in the presence of such structural shocks (Baumeister & Kilian, 2014b). This is a general problem that has many real life applications. For example, one may be interested in forecasting the stock price of a company tomorrow after hearing terrible or great news about the company after trading has concluded for the day. Companies may be interested in forecasting the demand of their products to adjust production after they were involved in a brand crisis, but they only have recent sales data for which the company is operating well. All is not lost in this forecasting setting, one may be able to supplement the present forecast with past data borrowed from other time series which contain similar structural shocks. The core idea of our methodology is to sensibly aggregate similar

past realized shock effects which arose from other time series, and then incorporate the aggregated shock effect estimator into the present forecast. Our method of combining shock effects embraces ideas from conditional forecasting (Baumeister & Kilian, 2014b; Kilian & Lütkepohl, 2017), time series pooling using cross-sectional panel data (Ramaswamy, DeSarbo, Reibstein & Robinson, 1993; Pesaran, Shin & Smith, 1999; Hoogstrate, Palm & Pfann, 2000; Baltagi, 2008; Koop & Korobilis, 2012; Liu, Moon & Schorfheide, 2020), forecasting with judgement and models (Svensson, 2005; Monti, 2008), synthetic control methodology (Abadie, Diamond & Hainmueller, 2010; Agarwal, Alomar, Sarker, Shah, Shen & Yang, 2020), and expectation shocks (Croushore & Evans, 2006; Baumeister & Kilian, 2014a; Clements, Galvão et al., 2019).

We study the post-shock forecasting problem in the context of additive shock effects in linear autoregressive models. In this 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. In this setting, we treat the additive shock effect as an observed random effect, obtained from a distribution of similar shocks, that is parameterized as a conditional fixed effect within an autoregressive model. The shock effect is then estimated using ordinary least squares (OLS). The researcher must move beyond the modeling paradigm that they were previously working under to accommodate this new shock effect (Monti, 2008; Svensson, 2005). One method for estimating the shock effect is to produce a conditional point forecast where a sequence of non-zero future structural shocks are conditioned upon and estimated (Baumeister & Kilian, 2014b). Such conditional point forecasts are appropriate when the shock sequence considered is within the range of historical experience (Kilian & Lütkepohl, 2017). On the other hand, our methodology allows for the inclusion of outside data sources and covariates into this conditional forecasting framework, provided that the shock effects from outside data sources arose from the same data generating process as the shock effect under study. For further differences of assumptions on shocks, our methodology allows for unprecedented shocks and no observation of shocks from the recent past. Our shock effect setting is also similar to the data generating mechanism that Tsay (1986) studied for ARMA models. Tsay (1986) studied additive outliers that contaminates the data at only one or multiple specific points, and innovational outliers that impact the subsequent responses. Please clarify:... Our shock effect of interest is similar to innovational outliers in the sense that it is an additive effect in the model of response only at a particular time point but it impacts the subsequent

responses through an autoregressive structure.

In our methodological framework, the researcher creates a synthetic panel of 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) (Abadie et al., 2010). As in SCM, care is needed when forming the donor pool of time series. However, there are key differences between our framework and SCM. We assume that the time series forming the donor pool are independent from the time series under study before the timing of the shock. Moreover, the shocks from candidate time series in the donor pool, together with the shock in the time series of interest, are assumed to be from a common family of distributions with existing second moments.

We then estimate the shock effects that are present in the time series forming the donor pool for which post-shock responses are observed. These estimated shock effects are then aggregated and this aggregated estimator is taken to be the estimator for the shock effect in the time series of interest. Finally, this aggreagted estimator is then added to a forecast for the yet to be realized post-shock response of the time series of interest. Our adjustment is similar to various adjustment methodologies suggested by Clements & Hendry (1996) that use the residual at end of observed time horizon for h-ahead forecasts. However, the approach of Clements & Hendry (1996) is not robust to unprecedented and unobserved shock since the shock signal is not embedded in the data. Under a similar setting that studies equilibrium shifts, Castle, Clements & Hendry (2015) propose robust general approaches for forecasting in vector equilibrium-correction models. However, their forecasting methodology requires conditioning on the past data that contain shock signals. This approach is therefore not applicable in the post-shock setting.

The shock effects in our post-shock forecasting framework are similar to "expectation shocks," studied in Clements et al. (2019). Clements et al. (2019) allows for consistent estimation of expectation shocks under a vector autoregressive model, possibly involving an instrumental variable approach, as in Croushore & Evans (2006). In our context, the yet to be observed shock effect of interest arises as a random effect whose assumed generative distribution can only be partially estimated. This partial estimation only uses shock effect estimates from the time series forming the donor pool.

In this article, we will assume a simple autoregressive data generating process with a general random-effect structure. These specifications are similar to the "K latent pooling" framework of

Ramaswamy et al. (1993). However, our model formulation is more general than Ramaswamy et al. (1993). In our model, the donor pool can consist of dependent time series, but the time series forming the donor pool are required to be independent from the time series of interest. However, mutual independence among time series in the donor pool can aid prospective evaluation of the reliability of our method. We consider three aggregation techniques: simple averaging, inverse-variance weighted averaging, and similarity weighting. The latter technique is similar to the weighting in synthetic control methodology (Abadie et al., 2010). Our autoregressive model will consider present day covariates to better motivate similarity weighting. The considered adjustment strategies all target the mean of the shock effect distribution. 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 can be estimated by a residual bootstrap procedure that we develop. 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 with two real data analyses. The first of these focuses on forecasting the stock price of Conoco Phillips after March 9th, 2020, in the midst of the COVID-19 pandemic and an oil supply war. The second of these focuses on forecasting the stock price of Apple Inc. after the company releases their latest installment of the MacBook on November 10th, 2020. We will show that our proposed adjustment estimators yield much better results than no adjustment in these settings. The example of Conoco Phillips provides an empirical application of our methods in unprecedented shocks whereas the Apple example provides a case for post-shock prediction where the shock has precedent. We now motivate our framework for post-shock prediction.

2. Setting

We will suppose that a researcher has time series data $(y_{i,t}, \mathbf{x}_{i,t})$, for $t = 1, \ldots, T_i$ and $i = 1, \ldots, n+1$, where $y_{i,t}$ is a scalar response and $\mathbf{x}_{i,t}$ is a vector of covariates that are revealed to
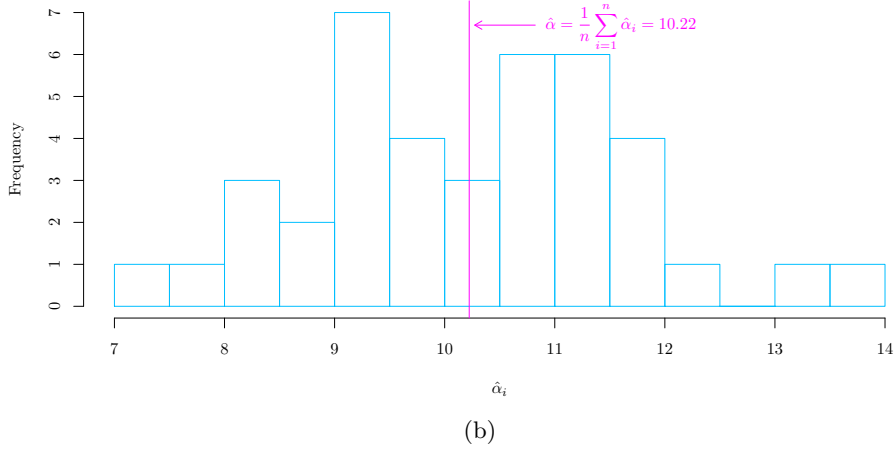
Figure 1: The time series experience a shock at $T_1^* + 1 = 126$ with true shock effect $\alpha = 9.21$. (a) presents the comparison of the prediction without adjustment and one that uses simple averaging of estimated shock effects given a donor pool of $n = 40$ time series. (b) shows the histogram for the least square estimates $\hat{\alpha}_i$s used in estimating shock α_1 for $i = 2, \ldots, 41$. The magenta dots represent $\hat{\alpha}_i$ from the time series forming the donor pool. $\hat{y}_{T_1^*+1}^1$ is the prediction without scalar adjustment of $\hat{\alpha}$ whereas $\hat{y}_{T_1^*+1}^2$ is the one with. In this example, they differ only by an adjustment $\hat{\alpha} = 10.22$. It is clear that $\hat{y}_{T_1^*+1}^2$ performs better than $\hat{y}_{T_1^*+1}^1$.

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 MSE,

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

5

and root mean squared error (RMSE), given by $\sqrt{R_T}$, in our analyses. In this article, we focus on post-shock prediction where forecasting methods only differ at the next future time point. Thus the MSE reduces to the magnitude $\mathrm{E}(\hat{y}_{1,t} - y_{1,t})^2$.

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 $(y_{i,t}, \mathbf{x}_{i,t})$, for $t = 1, \ldots, T_i$ and $i = 1, \ldots, n+1$. In this article, we consider a special case that the model includes random effects for the covariates, autoregressive structure, and an additional shock effect, whose presence or absence is given by the indicator variable $D_{i,t}$. We will assume that the donor pool time series are independent of the time series under study. The details of this model are in the next section.

Figure 1 provides 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 after T_1^*, but does have observations from a donor pool of forty 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 estimation procedure that ignores the shock. It is apparent from Figure 1 that adjusting forecasts in this manner 1) leads to a reduction in forecasting risk; 2) does not fully recover the true shock effect. We evaluate the performance of this post-shock forecasting methodology throughout this article; we outline situations for when it is expected to work and when it is not.

2.1. Model setup

We now describe the assumed autoregressive models with random effects for which post-shock aggregated estimators are provided. The basic structures of these models are the same for all time-series in the analysis, the differences between them lie in the setup of the shock effect distribution.

Let $I(\cdot)$ be an indicator function, T_i be the time length of the time series i for $i = 1, \ldots, n+1$, and T_i^* be the time point just before the one when the shock is known to occur, with $T_i^* < T_i$. For

$t = 1, \ldots, T_i$ and $i = 1, \ldots, n + 1$, the model $\mathcal{M}_1$ is defined as

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

where $D_{i,t} = I(t = T_i^* + 1)$ and $\mathbf{x}_{i,t} \in \mathbb{R}^p$ with $p \geq 1$. We assume that the $\mathbf{x}_{i,t}$s are fixed. Let $|x|$ denote the absolute value of x for $x \in \mathbb{R}$. For $i = 1, \ldots, n + 1$ and $t = 1, \ldots, T_i$, the random effects structure for $\mathcal{M}_1$ is:

$$\eta_i \overset{iid}{\sim} \mathcal{F}_\eta \text{ with } \mathrm{E}_{\mathcal{F}_\eta}(\eta_i) = 0, \mathrm{Var}_{\mathcal{F}_\eta}(\eta_i) = \sigma_\eta^2$$

$$\phi_i \overset{iid}{\sim} \mathcal{F}_\phi \text{ where } |\mathcal{F}_\phi| < 1,$$

$$\theta_i \overset{iid}{\sim} \mathcal{F}_\theta \text{ with } \mathrm{E}_{\mathcal{F}_\theta}(\theta_i) = \mu_\theta, \mathrm{Var}_{\mathcal{F}_\theta}(\theta_i) = \Sigma_\theta^2$$

$$\alpha_i \overset{iid}{\sim} \mathcal{F}_\alpha \text{ with } \mathrm{E}_{\mathcal{F}_\alpha}(\alpha_i) = \mu_\alpha, \mathrm{Var}_{\mathcal{F}_\alpha}(\alpha_i) = \sigma_\alpha^2$$

$$\varepsilon_{i,t} \overset{iid}{\sim} \mathcal{F}_\varepsilon \text{ with } \mathrm{E}_{\mathcal{F}_\varepsilon}(\varepsilon_{i,t}) = 0, \mathrm{Var}_{\mathcal{F}_\varepsilon}(\varepsilon_{i,t}) = \sigma^2,$$

$$\eta_i \perp\!\!\!\perp \alpha_i \perp\!\!\!\perp \phi_i \perp\!\!\!\perp \theta_i \perp\!\!\!\perp \varepsilon_{i,t},$$

where $\sigma > 0$. 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 with an additional additive mean-zero error. For $i = 1, \ldots, n + 1$, the random effects structure for this model (model $\mathcal{M}_2$) is:

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

where the added random effects are

$$\tilde{\varepsilon}_i \overset{iid}{\sim} \mathcal{F}_{\tilde{\varepsilon}} \text{ with } \mathrm{E}_{\mathcal{F}_{\tilde{\varepsilon}}}(\tilde{\varepsilon}_i) = 0, \mathrm{Var}_{\mathcal{F}_{\tilde{\varepsilon}}}(\tilde{\varepsilon}_i) = \sigma_\alpha^2$$

$$\eta_i \perp\!\!\!\perp \alpha_i \perp\!\!\!\perp \phi_i \perp\!\!\!\perp \theta_i \perp\!\!\!\perp \varepsilon_{i,t} \perp\!\!\!\perp \tilde{\varepsilon}_i.$$

where $\sigma_a > 0$. We further define $\tilde{\alpha}_i = \mu_\alpha + \delta_i' \mathbf{x}_{i,T_i^*+1}$. We will investigate the post-shock aggregated estimators in $\mathcal{M}_2$ in settings where δ_i is either fixed or random. We let $\mathcal{M}_{21}$ denote model $\mathcal{M}_2$ with $\delta_i = \delta$ for $i = 1, \ldots, n + 1$, where δ is a fixed unknown parameter. We let $\mathcal{M}_{22}$ denote model $\mathcal{M}_2$ with the following random effects :

$$\delta_i \overset{iid}{\sim} \mathcal{F}_\delta \text{ with } \mathrm{E}_{\mathcal{F}_\delta}(\delta_i) = \mu_\delta, \mathrm{Var}_{\mathcal{F}_\delta}(\delta_i) = \Sigma_\delta$$

$$\delta_i \perp\!\!\!\perp \tilde{\varepsilon}_i.$$

We further define the parameter sets

$$\Theta = \{(\eta_i, \phi_i, \theta_i, \alpha_i, \mathbf{x}_{i,t}, y_{i,t-1}, \delta_i) \colon t = 1, \ldots, T_i, i = 2, \ldots, n+1\}$$
$$\Theta_1 = \{(\eta_i, \phi_i, \theta_i, \alpha_i, \mathbf{x}_{i,t}, y_{i,t-1}, \delta_i) \colon t = 1, \ldots, T_i, i = 1\}$$

(3)

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

2.2. Forecast

In this section we show how post-shock aggregate estimators improve upon standard forecasts that do not account for the shock effect.

More formally, we will consider the following candidate forecasts:

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

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

where $\hat{\eta}_1$, $\hat{\phi}_1$, and $\hat{\theta}_1$ are all OLS estimators of η_1, ϕ_1, and θ_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.

Throughout the rest of this article we highlight when the information from the time series donor pool, indexed by $\{y_{i,t} \colon t = 1, \ldots, T_i, i = 2, \ldots, n+1\}$, can be used to construct a shock effect estimator $\hat{\alpha}$ in which Forecast 2 beats Forecast 1 and vice-versa. We will consider the different dynamic panel models $\mathcal{M}_1$, $\mathcal{M}_{21}$, and $\mathcal{M}_{22}$. Improvement will be measured by assessing the reduction in risk that Forecast 2 offers over Forecast 1. We outline the theoretical details of risk-reduction in Section 3. We specifically focus on predictions for y_{1,T_1^*+1}, the first post-shock response. It is important to note that, in general, $\hat{\alpha}$ does not converge to α_1 in any sense. Despite this shortcoming, 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, and for several estimation techniques.

2.3. Shock effect estimators

We now construct the aggregated estimators of the shock effects that appear in Forecast 2 (see Section 2.2). 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 2.1). For $i = 2, \ldots, n+1$, all parameters in this model will be estimated via ordinary least squares

(OLS) using historical data indexed by $t = 1, \ldots, T_i$. For $i = 1$, we estimate all the parameters, excluding α_1, using OLS for time periods indexed by $t = 1, \ldots, T_1^*$. Let $\hat{\alpha}_i$ be the OLS estimate of α_i, $i = 2, \ldots, n+1$. Note that parameter estimation under $\mathcal{M}_1$ is the same as that for $\mathcal{M}_{21}$ or $\mathcal{M}_{22}$. We emphasize that the α_is are realizations from a random effects distribution, and that the OLS estimators $\hat{\alpha}_i$ assumes that these realized values are conditional fixed effects in an autoregressive model, as in (1) and (2).

The candidate estimators for α_1 are now introduced. Define the adjustment estimator for time series $i = 1$ by

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

We can use $\hat{\alpha}_{\mathrm{adj}}$ as an estimator for the unknown α_1 term for which no meaningful estimation information otherwise exists. It is intuitive that $\hat{\alpha}_{\mathrm{adj}}$ should perform well under $\mathcal{M}_1$ where we assume that α_is share the same mean for $i = 1, \ldots, n+1$. However, it can also be shown that $\hat{\alpha}_{\mathrm{adj}}$ may be less favorable in $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$, which will be discussed in detail in Section 3.

We also consider the inverse-variance weighted estimator in practical settings where the T_is and T_i^*s vary greatly across $i = 2, \ldots, n+1$. The inverse-variance weighted estimator is defined as

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

where $\hat{\alpha}_i$ is the OLS estimator of α_i, $\hat{\sigma}_i$ is the residual standard error from OLS estimation, and $\mathbf{M}_i$ 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, $\hat{\alpha}_{\mathrm{IVW}}$ can be a reasonable estimator in practical settings, especially when the T_i^*s very considerably. We do not provide closed form expressions for $\mathrm{E}(\hat{\alpha}_{\mathrm{IVW}})$ and $\mathrm{Var}(\hat{\alpha}_{\mathrm{IVW}})$ but empirical performance of $\hat{\alpha}_{\mathrm{IVW}}$ is assessed via Monte Carlo simulation (see Section 4).

We now motivate a weighted-adjustment estimator appropriate for model $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$. Our weighted-adjustment estimator is inspired by the weighting techniques in synthetic control methodology (SCM) developed in Abadie et al. (2010). However, our weighted-adjustment estimator is not a pure causal estimator, and our estimation premise is a reversal of the SCM context in Abadie et al. (2010). In our context, we predict the unobserved post-shock response y_{1,T_1^*+1} (intervention is absent) using post-shock information from other time series whose post-shock responses are all observed (intervention is present).

9

We use similar notation as that in Abadie et al. (2010) to describe 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

$$\mathbf{X}_1 = \mathbf{x}'_{1,T_1^*+1}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}'_{2,T_2^*+1} \\ \vdots \\ \mathbf{x}'_{n+1,T_{n+1}^*+1} \end{pmatrix},$$

$$\hat{\mathbf{X}}_1(\mathbf{W}) = \mathbf{W}'\mathbf{X},$$

where $\mathbf{X}_1, \hat{\mathbf{X}}_1(\mathbf{W}) \in \mathbb{R}^{1 \times p}$. Define $\mathcal{W} = \{\mathbf{W} \in [0,1]^n : 1'_n \mathbf{W} = 1\}$. Suppose there exists $\mathbf{W}^* \in \mathcal{W}$ with $\mathbf{W}^* = (w_2^*, \ldots, w_{n+1}^*)'$ such that

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

Note that (5) tries to find $\mathbf{W}^*$ such that $\mathbf{x}_{1,T_1^*+1}$ is a convex combination of $\mathbf{x}_{i,T_i^*+1}$ for $i = 2, \ldots, n+1$ with weights $\mathbf{W}^*$. Therefore, $\mathbf{W}^*$ should exist as long as $\mathbf{X}_1$ falls in the convex hull of $\left\{\mathbf{x}'_{2,T_2^*+1}, \ldots, \mathbf{x}'_{n+1,T_{n+1}^*+1}\right\}$. Our weighted-adjustment estimator will therefore perform well when the pool of time series posses similar covariates to the time series for which no post-shock responses are observed. We compute $\mathbf{W}^*$ as

$$\mathbf{W}^* = \arg\min_{\mathbf{W} \in \mathcal{W}} \left\| \mathbf{X}_1 - \hat{\mathbf{X}}_1(\mathbf{W}) \right\|_p. \tag{6}$$

Abadie et al. (2010) commented that we can select $\mathbf{W}^*$ in (6) so that (5) holds approximately and that weighted-adjustment estimation techniques of this form are not appropriate when the fit is poor. Note that $\mathbf{W}^*$ is not random since the covariates are assumed to be fixed. Since $\mathcal{W}$ is a closed and bounded subset of $\mathbb{R}^n$, $\mathcal{W}$ is compact. Because the objective function is continuous in $\mathbf{W}$, $\mathbf{W}^*$ will always exist. The weighted-adjustment estimator for the shock effect α_1 obtained from the weights $\mathbf{W}^*$ is defined as

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

and we further define $\mathbf{V} = (\mathbf{x}_{2,T_2^*+1}, \ldots, \mathbf{x}_{n+1,T_{n+1}^*+1}) \in \mathbb{R}^{p \times n}$.

Proposition 1. *If $\mathbf{V}$ has full rank then there exists some $\mathbf{W}^*$ that solves (5) and this solution is unique.*

Proposition 1 details some conditions when $\mathbf{W}^*$ is unique. If $\mathbf{V}$ is full rank and the true solution lies in the convex and compact $\mathcal{W}$, a sufficient condition for $\mathbf{W}^*$ to be unique is when $p \geq n$ in our setting. Non-uniqueness might be a concern theoretically. This is due to the fact that infinitely many different weights can lead to infinitely many non-unique $\hat{\alpha}_{\text{wadj}}$s, all targeting α_1. For example, consider the case when $n = 2$, we have $\text{Var}(\hat{\alpha}_2) = 1$ and $\text{Var}(\hat{\alpha}_3) = 2$, the covariates are such that $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}_3$, and we consider $\mathbf{W}_1^* = (1, 0)$ and $\mathbf{W}_2^* = (0, 1)$ as solutions to (6). In this hypothetical, the weighted adjustment estimator constructed from $\mathbf{W}_1^*$ has variance 1 whereas the same quantity constructed from $\mathbf{W}_2^*$ has variance 2. It is possible to resolve this issue by selecting a unique weight $\mathbf{W}^*$ that optimizes a desirable objective function, prior to which one should find the bases spanning the subspace of $\mathbf{W}^*$ satisfying (5). Empirically, we find some simulation evidence that non-uniqueness of $\mathbf{W}^*$ is not problematic for inferences in our setting. Specifics can be seen in Section 4 of the Supplementary Materials. We close this Section with some remarks about our shock-effect estimators:

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

Remark 2. Our forecasting premise and estimation construction shares similarities with Bayesian viewpoints. From a Bayesian perspective, if we assign a prior π to α_1, $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ can be interpreted as the Bayes rules with respect to π under different loss functions. If the sampling distribution of the data and π are known, it is possible to compute the Bayes risks of $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ with respect to π, thus enabling comparisons among them. Additionally, from Theorem 2.4 in Chapter 5 of Lehmann & Casella (2006), $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ are admissible if they are unique with probability one.

Remark 3. We also provide some recommended choices for shock-effect estimators in different situations. First, $\hat{\alpha}_{\text{adj}}$ is usually recommended when we believe the shock is i.i.d from a distribution. Second, $\hat{\alpha}_{\text{IVW}}$ is recommended when the time length T_i in the donor pool varies vastly across the donor pool. Third, $\hat{\alpha}_{\text{wadj}}$ is recommended when we assume the shocks are from the same family of distribution, and the mean of the shock depends on some covariates.

3. Forecast risk, risk-reduction, and properties of shock effects estimators

In this section, we discuss the properties that are related to forecast risk-reduction. It is useful to first 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 follows:

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

(ii) Under $\mathcal{M}_{21}$, $\text{E}(\hat{\alpha}_{\text{adj}}) = \mu_\alpha + \frac{1}{n}\sum_{i=2}^{n+1} \delta' \mathbf{x}_{i,T_i^*+1}$ and $\text{E}(\hat{\alpha}_{\text{wadj}}) = \mu_\alpha + \delta' \mathbf{x}_{1,T_1^*+1}$.

(iii) Under $\mathcal{M}_{22}$, $\text{E}(\hat{\alpha}_{\text{adj}}) = \mu_\alpha + \frac{1}{n}\sum_{i=2}^{n+1} \mu_\delta' \mathbf{x}_{i,T_i^*+1}$ and $\text{E}(\hat{\alpha}_{\text{wadj}}) = \mu_\alpha + \mu_\delta' \mathbf{x}_{1,T_1^*+1}$.

Formal justification for these results can be found in the Appendix. Note that $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ are not unbiased estimators for α_1. However, under $\mathcal{M}_1$, $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$ are unbiased estimators for $\text{E}(\alpha_1) = \mu_\alpha$. Nevertheless, $\hat{\alpha}_{\text{adj}}$ is a biased estimator for $\text{E}(\alpha_1)$. On the other hand, $\hat{\alpha}_{\text{wadj}}$ is an unbiased estimator for $\text{E}(\alpha_1)$ under both $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$. We collect these results in the following proposition:

Proposition 2.

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

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

Unbiasedness properties for $\text{E}(\alpha_1)$ of $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{wadj}}$ allow for simple conditions for risk-reduction to hold, and more importantly motivates a bootstrap estimation procedure for evaluation of these conditions, discussed in Section 3.1 and 3.2 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 $\mathcal{M}_{21}$,

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

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

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

$$\mathrm{Var}(\hat{\alpha}_{\mathrm{adj}}) = \frac{\sigma^2}{n^2}\sum_{i=2}^{n+1}\mathrm{E}\{(\mathbf{M}_i'\mathbf{M}_i)_{22}^{-1}\} + \frac{1}{n^2}(\mathbf{x}_{i,T_i^*+1}'\Sigma_\delta\mathbf{x}_{i,T_i^*+1} + \sigma_\alpha^2),$$

$$\mathrm{Var}(\hat{\alpha}_{\mathrm{wadj}}) = \sigma^2\sum_{i=2}^{n+1}(w_i^*)^2\mathrm{E}\{(\mathbf{M}_i'\mathbf{M}_i)_{22}^{-1}\} + \sum_{i=2}^{n+1}(w_i^*)^2(\mathbf{x}_{i,T_i^*+1}'\Sigma_\delta\mathbf{x}_{i,T_i^*+1} + \sigma_\alpha^2).$$

Formal justification for these results can be found in Section 8 of the Appendix. Note that the variances are not comparable in closed-form because of the term $\mathrm{E}\{(\mathbf{M}_i'\mathbf{M}_i)_{22}^{-1}\}$, where $\mathbf{M}_i$ is the design matrix for OLS with respect to time series for $i = 2,\ldots,n+1$ from Section 2.3. This term exists because of the inclusion of the random lagged response in our autoregressive model formulation.

Section 3.1 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 conditions in practice, estimation is required. Sections 3.2 introduce a residual bootstrap procedure which estimates the involved parameters in those conditions and thus motivates prospective decision-making about whether $\hat{\alpha}_i$ reduces the risk. Section 3.3 describes our leave-one-out cross validation procedures, which prospectively estimate the correctness of such decision without observation of the post-shock response for the time series under study. Our simulations verify 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 $\mathcal{M}_{22}$. For an adjustment estimator $\hat{\alpha}$, we will write the risk-reduction as $\Delta(\hat{\alpha}) = R_{T_1^*+1,1} - R_{T_1^*+1,2}$ where $R_{T_1^*+1,2}$ is the risk of Forecast 2 calculated using the adjustment estimator $\hat{\alpha}$.

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

Recall that Proposition 2 implies that the adjustment estimator $\hat{\alpha}_{\mathrm{adj}}$ and weighted-adjustment estimator $\hat{\alpha}_{\mathrm{wadj}}$ are unbiased for $\mathrm{E}(\alpha_1)$ under $\mathcal{M}_1$. With this result, we will have the following proposition that specify the conditions that are necessary for risk reduction:

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

(i) $\Delta(\hat{\alpha}_{\text{adj}}) > 0$ *when* $\text{Var}(\hat{\alpha}_{\text{adj}}) < \mu_\alpha^2$.

(ii) if $\mathbf{W}^*$ *satisfies (5), then* $\Delta(\hat{\alpha}_{\text{wadj}}) > 0$ *when* $\text{Var}(\hat{\alpha}_{\text{wadj}}) < \mu_\alpha^2$.

Proposition 3 says 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. In this setting, under $\mathcal{M}_1$, $\Delta(\hat{\alpha}_{\text{adj}}) = \mu_\alpha^2 - \text{Var}(\hat{\alpha}_{\text{adj}})$ and $\Delta(\hat{\alpha}_{\text{wadj}}) = \mu_\alpha^2 - \text{Var}(\hat{\alpha}_{\text{wadj}})$. From Proposition 3, we obtain a risk-reduction condition

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

Condition (7) implies two facts: (1) adjustment (Forecast 2) is preferable to no adjustment (Forecast 1) asymptotically in n whenever $\mu_\alpha \neq 0$; (2) In the finite donor pool settings, adjustment is preferable to no adjustment when μ_α is large relative to its variability and overall regression variability.

If $\mathbf{W}^*$ does not satisfy (5), its unbiased properties for $\text{E}(\alpha_1)$ should hold approximately when the fit in (6) is appropriate as commented in Section 2.3. From Proposition 3 and the variance expression for $\hat{\alpha}_{\text{wadj}}$, the risk-reduction condition for $\hat{\alpha}_{\text{wadj}}$ is

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

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. Sections 3.2 and 3.3 provide detailed treatments on how to estimate the sign of $\Delta(\hat{\alpha})$ 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 2 implies that $\hat{\alpha}_{\text{wadj}}$ is an unbiased estimator of $\text{E}(\alpha_1)$. We now state conditions for risk-reduction.

Proposition 4. *If* $\mathbf{W}^*$ *satisfies (5), then* $\Delta(\hat{\alpha}_{\text{wadj}}) > 0$ *when* $\text{Var}(\hat{\alpha}_{\text{wadj}}) < (\text{E}(\alpha_1))^2$ *under* $\mathcal{M}_{21}$ *and* $\mathcal{M}_{22}$.

Under Proposition 4, we can obtain a risk-reduction inequality that is similar to (8),

$$\mathrm{Var}(\hat{\alpha}_{\mathrm{wadj}}) = \sigma^2 \sum_{i=2}^{n+1} (w_i^*)^2 \mathrm{E}\{(\mathbf{M}_i'\mathbf{M}_i)_{22}^{-1}\} + \sum_{i=2}^{n+1} (w_i^*)^2 (\mathbf{x}_{i,T_i^*+1}'\Sigma_\delta\mathbf{x}_{i,T_i^*+1} + \sigma_\alpha^2) < (\mathrm{E}(\alpha_1))^2,$$

where $\mathbf{x}_{i,T_i^*+1}'\Sigma_\delta\mathbf{x}_{i,T_i^*+1} + \sigma_\alpha^2$ may be replaced with σ_α^2 in $\mathcal{M}_{21}$. The conclusions and intuitions will be identically the same as what we have in Section 3.1.1. Proposition 2 shows that $\hat{\alpha}_{\mathrm{adj}}$ is a biased estimator of $\mathrm{E}(\alpha_1)$ under $\mathcal{M}_{21}$ and $\mathcal{M}_{22}$ in general. Hence, Proposition 3 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 1, we will achieve risk-reduction when

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

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 2.3, 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 $\mathrm{E}(\alpha_1)$. We can adapt the above to derive a risk-reduction condition for $\hat{\alpha}_{\mathrm{IVW}}$: under $\mathcal{M}_1$, $\mathcal{M}_{21}$, and $\mathcal{M}_{22}$, $\Delta(\hat{\alpha}_{\mathrm{IVW}}) > 0$ when $\mathrm{Var}(\hat{\alpha}_{\mathrm{IVW}}) + (\mathrm{E}(\hat{\alpha}_{\mathrm{IVW}}) - \mathrm{E}(\alpha_1))^2 < (\mathrm{E}(\alpha_1))^2$. In fact, more generally, using similar proof of Lemma 1, it can be shown that under $\mathcal{M}_2$, the risk-reduction quantities are

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

We estimate $\Delta(\hat{\alpha})$ using bootstrap and leave-one-out cross validation procedures developed in Sections 3.2 and 3.3.

3.2. Bootstrap for risk-reduction evaluation problems

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 Efron & Tibshirani (1986) and Chapter 12 of Kilian & Lütkepohl (2017). Our procedure involves the resampling of the residuals which are assumed to be the realizations of an iid process.

Our first bootstrap procedure is as follows: Let B be the bootstrap sample size. At iteration b, first resample the indices $I = \{2, \ldots, n+1\}$ of the donor pool with replacement to form $I^{(b)}$ with cardinality n, where we note that the elements of $I^{(b)}$ may not be unique in terms of their indices in the donor pool. Initialize $y_{i,0}^{(b)} = y_{i,0}$ for all $i \in I^{(b)}$. Then, resample the residuals under models $\mathcal{M}_1$ or $\mathcal{M}_2$, compute the bootstrapped response $y_{i,t}^{(b)}$ for $t \in \{1, \ldots, T_i\}$ using the model estimated by original data, and obtain shock effect estimators for each of the time series in the donor pool for all $i \in I^{(b)}$. These shock effect estimators are then used to construct any of the adjustment estimators $\hat{\alpha}_{\text{adj}}^{(b)}$, $\hat{\alpha}_{\text{wadj}}^{(b)}$, and $\hat{\alpha}_{\text{IVW}}^{(b)}$, for $b = 1, \ldots, B$. We can then estimate distributional quantities of our shock effect estimators under our considered models with the bootstrap samples $\hat{\alpha}_{\text{adj}}^{(b)}$, $\hat{\alpha}_{\text{wadj}}^{(b)}$, and $\hat{\alpha}_{\text{IVW}}^{(b)}$, 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 of the Supplementary Materials.

We will explicitly use these bootstrapped samples of shock effect estimators to check the risk-reduction conditions in Propositions 3 and 4. Recall that $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$ and $\hat{\alpha}_{\text{IVW}}$ are unbiased estimators of their expectations, and $\hat{\alpha}_{\text{wadj}}$ is an unbiased estimator of $\text{E}(\alpha_1)$ under $\mathcal{M}_1$ and $\mathcal{M}_2$ from Proposition 2. 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 $\Delta(\hat{\alpha}_{\text{adj}})$ under model $\mathcal{M}_{21}$ or $\mathcal{M}_{22}$ with

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

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

We reiterate the philosophical distinction 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 arises from parameters and idiosyncratic error. A double bootstrap procedure with similar steps 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.

We stress that 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 $i \in \{2, \ldots, n+1\}$ (El Karoui & Purdom, 2018); see alternatives for residual bootstrapping in linear models in El Karoui & Purdom (2018).

3.3. Leave-one-out cross validation

In this section, we motivate a leave-one-out cross validation (LOOCV) procedure within 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 Hastie, Tibshirani & Friedman (2009). Recall in Section 2.1 that we have data $\{(\mathbf{x}_{i,t}, y_{i,t}): i = 1, \ldots, n+1, t = 1, \ldots, T_i\}$, where $\{(\mathbf{x}_{1,t}, y_{1,t}): t = 1, \ldots, T_1\}$ 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 $\{(\mathbf{x}_{m+1,t}, y_{m+1,t}): t = 1, \ldots, T_{m+1}\}$ as the time series of interest, and construct a new donor pool $\{(\mathbf{x}_{i,t}, y_{i,t}): i \in \mathcal{I}_m, t = 1, \ldots, T_i\}$, 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.

LOOCV can be very computationally intensive when n is large, especially when combined with bootstrapping. To alleviate these concerns, one 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 $\{(\mathbf{x}_{m+1,t}, y_{m+1,t}): t = 1, \ldots, T_{m+1}\}$ as the time series of interest, and construct a new donor pool $\{(\mathbf{x}_{i,t}, y_{i,t}): i \in \mathcal{I}, t = 1, \ldots, T_i\}$, 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 3.2. 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 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 makes 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 makes 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 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 indices.

Remark 4. Note that we allow the time series within the donor pool to be dependent but donor pool should be independent of the time series of interest. However, if we assume the mutual independence structure, $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$ will be an almost unbiased estimator of $\mathrm{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ (Marden, 2015, Page 222).

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_δ approximating the design of $\mathcal{M}_{21}$. We consider $p = 25$ and $\mu_\alpha = 2$, where $p = 25$ is set to satisfy conditions in Proposition 1. The parameter setup of our simulations is detailed as follows:

$$\phi_i \overset{iid}{\sim} U(0,1),$$
$$\delta_i \overset{iid}{\sim} \mathcal{N}(1, 0.5),$$
$$\theta_i \overset{iid}{\sim} \mathcal{N}(0,1),$$
$$\alpha_i \overset{iid}{\sim} \mathcal{N}(\mu_\alpha, \sigma_\alpha^2),$$
$$T_i \overset{iid}{\sim} \min\left(90, [\text{Gamma}(15, 10)]\right),$$
$$T_i^* \overset{iid}{\sim} U(p + 4, T_i - 1),$$
$$\varepsilon_{i,t} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

where the $[\cdot]$ operation in the generation of the T_is corresponds to rounding the input to the nearest integer and the discrete uniform distribution is used to generate the T_i^*s. 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 $\hat\alpha_{\text{IVW}}$ that may perform well in time series with varying lengths. Additionally, we generated all of the entries of the covariate vectors to be realizations from a Gamma$(1, 2)$ distribution.

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 our computations. Means and standard errors for estimated quantities will be recorded. Our LOOCV procedure will consider $k = 5$ random draws. Recall in Section 3.3 that B 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 (10) and (11) to assess correct decision making. We will also consider the Euclidean distance between the post-shock forecasts $\hat{y}_{1,T_1^*+1}$, $\hat{y}_{1,T_1^*+1} + \hat{\alpha}_{\text{adj}}$, $\hat{y}_{1,T_1^*+1} + \hat{\alpha}_{\text{wadj}}$, and $\hat{y}_{1,T_1^*+1} + \hat{\alpha}_{\text{IVW}}$ 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

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 2.1) for $\mathcal{B}_u$ and $\mathcal{B}_f$ (see Section 3.2) 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 $\text{E}(\text{E}(\alpha_1)) = 52$, 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 1 in the Appendix, Section 8.4.

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 ranges from strong to nearly nonexistent. Results are in Table 2 in the Appendix, Section 8.4.

First, assuming that $\bar{C}^{(k)}(\delta_{\hat{\alpha}})$ well estimates $\text{E}(\mathcal{C}(\delta_{\hat{\alpha}}))$ and fixing n, we observe from Table 1 that the decision making using $\delta_{\hat{\alpha}}$ is nearly correct for $\hat{\alpha} \in \mathcal{A}$ when σ_α is small. The reasons can be explained as follows: When σ_α is small, the signal of the covariates is strong so that $\hat{\alpha}_{\text{wadj}}$ will be expected to capture the signal according to construction of $\hat{\alpha}_{\text{wadj}}$ in Section 2.3. Moreover, when σ_α is small, $\mathcal{M}_{22}$ and $\mathcal{M}_{21}$ are very similar and estimation of $\text{E}(\alpha_1)$ should be nearly unbiased according to Proposition 2. However, when the signal of the covariates is poor (σ_α is large), the decision rule $\delta_{\hat{\alpha}}$ becomes unreliable for $\hat{\alpha} \in \mathcal{A}$. This is to be expected since the bootstrap estimates become more biased. However, users are warned by $\bar{C}^{(k)}(\delta_{\hat{\alpha}})$ and they have an idea of the effectiveness of $\delta_{\hat{\alpha}}$. Second, fixing σ_α, we can observe that the rate in which we make correct decisions, $\delta_{\hat{\alpha}}$, increases when n increases.

Additionally, we observe that in most cases $\delta_{\hat{\alpha}_{\text{wadj}}}$ reports that $\hat{\alpha}_{\text{wadj}}$ reduces the risk even when $\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{wadj}}})$ starts to break down. Recall from Section 3.2 that $\hat{\Delta}(\hat{\alpha})$ contains the squared bias for estimating $\text{E}(\alpha_1)$. But it is not present for $\hat{\Delta}(\hat{\alpha}_{\text{wadj}})$ since we applied the fact $\hat{\alpha}_{\text{wadj}}$ is unbiased for $\text{E}(\alpha_1)$ from Proposition 2 in plugging it in with replacing $\text{E}(\alpha_1)$. Therefore, when the signal from covariates is poorer, $\delta_{\hat{\alpha}_{\text{wadj}}}$ becomes less conservative. Besides, the averaged $I\big(\hat{\Delta}(\hat{\alpha}) > 0\big)$ times $\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}})$ can provide an approximation for the probability that $\hat{\alpha}$ actually reduces the risk assuming an symmetry of 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 $\hat{\alpha}_{\text{adj}}$ reduces the risk is approximately $0.83 \times 0.59 = 0.490$ from Table 1. In other words, the probability that $\hat{\alpha}$ reduces the risk has the same pattern as $\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}})$ has with n and σ_α for $\hat{\alpha} \in \mathcal{A}$.

In the last four columns of Table 1 we can see that as σ_α increases, the prediction appears to be poorer. When $\sigma_\alpha = 5, 10, 25$, forecasts using $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ are always better than the original forecast significantly. This does not hold generally for the case when $\sigma_\alpha = 50, 100$. It is reasonable that when the σ_α is large, it is difficult to find a reliable estimate of α_1. Nevertheless, no Monte-Carlo evidence has been found to support the claim that n plays a large role prediction.

From Table 2, we observe that as σ_α increases with σ fixed, $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$ decreases. Furthermore, as σ increases we have that $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$ decreases as well. Note that decision making ability hinges on the estimation of the parameters. Since $\hat{\alpha}_{\text{wadj}}$ is a linear combination of OLS estimates, as σ increases, $\text{Var}(\hat{\alpha}_{\text{wadj}})$ increases as well. Therefore, $\hat{\alpha}_{\text{wadj}}$ become more volatile and its estimation of $\text{E}(\alpha_1)$ can be less reliable. Those reasons can explain why an increase of σ_α contributes to a decrease of $\bar{\mathcal{C}}(\delta_{\hat{\alpha}})$. 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 $\hat{\alpha} \in \mathcal{A}$.

The averaged $I(\hat{\Delta}(\hat{\alpha}) > 0)$ (i.e., the guess) starts to decrease as σ increases. This is reasonable if we believe the bootstrap estimate $S_{\hat{\alpha}}^2$ provides a good approximation for $\text{Var}(\hat{\alpha})$ for $\hat{\alpha} \in \mathcal{A}$. The reasons are outlined as follows: Recall in Section 3.1.2, the conditions of risk-reduction propositions involve $(\text{E}(\alpha_1))^2 > \text{Var}(\hat{\alpha}) + (\text{E}(\hat{\alpha}) - \text{E}(\alpha_1))^2$ for $\hat{\alpha} \in \mathcal{A}$. Notice that $\text{Var}(\hat{\alpha})$ 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 $\text{Var}(\hat{\alpha})$ increases.

21

Simulation for $\mathcal{B}_f$ with the same parameter setup as that of $\mathcal{B}_u$ are implemented. See Table 3 and Table 4 for results in the Appendix, Section 8.4. From Table 1 and Table 3 we have that when n is moderately small ($n = 10$) and σ_α is small ($\sigma_\alpha = 5$), $\mathcal{B}_u$ is preferable to $\mathcal{B}_f$. For other situations, $\mathcal{B}_u$ and $\mathcal{B}_f$ perform similarly. It is likely that the extra randomness from sampling the donor pool with replacement compensates for the possible noises from a small donor pool. Concerning Table 2 and Table 4, 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 moderately small; otherwise, they are similar.

Simulation results corresponding to model $\mathcal{M}_1$ are listed in Section 3 in the Supplementary Materials. Results under model $\mathcal{M}_1$ are very similar to those of $\mathcal{M}_2$. The results show that: (1) the performance of $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{IVW}}$ are nearly the same and (2) in many situations, $\hat{\alpha}_{\text{adj}}$ and $\hat{\alpha}_{\text{IVW}}$ are better than $\hat{\alpha}_{\text{wadj}}$; 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.

In these simulations we have implicitly assumed that $\mathbf{W}^*$ is non-degenerate in the population. Recall that in Section 2.3 we noted that if there exists some $\mathbf{W}^*$ which satisfies (5) and $p < n$, then there will be infinitely many solutions $\mathbf{W}^*$. In this scenario, $\mathbf{W}^*$ will take values on the boundary of $\mathcal{W}$, in which case bootstrapping may fail to estimate the distribution of $\hat{\alpha}_{\text{wadj}}$ (Andrews, 2000). When $p < n$ and there exists some $\mathbf{W}^* \in \mathcal{W}$ that satisfies (5), $\mathcal{B}_u$ fails since the non-uniqueness due to $p < n$ 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 $\mathcal{W}$. Empirically, simulations in Section 4 of the Supplementary Materials show that non-uniqueness does not seriously compromise inferences under $\mathcal{B}_u$.

Our simulations provide some empirical evidence that $\mathcal{B}_u$ is preferred when the donor pool size is moderately small ($n < 10$) and the signal from the covariates is believed to be large. If $p < n$ and

non-uniqueness problem is of concern for theoretical reasons, $\mathcal{B}_f$ may be preferred. In most cases, the choice between $\mathcal{B}_u$ and $\mathcal{B}_f$ would be up to personal preferences related to the philosophical differences between these bootstrap procedures.

5. Empirical examples

In this section, we discuss two real data examples of post-shock prediction. Section 5.1 discusses the prediction of Conoco Phillips stock occurring on March 9th, 2020. This shock relates to an oil price war and the COVID-19 pandemic. Section 5.2 applies our proposed method to the forecasting of Apple stock prices after releasing the latest version of their MacBook product. These examples provide empirical illustration of our methodology in settings where the shock is, respectively, unprecedented and with precedent. The adjusted forecast in Section 5.1 performs much better than the unadjusted original forecast, and the adjusted forecast in Section 5.2 is very close to the response after shock as reasonably expected.

5.1. Forecasting Conoco Phillips stock in the presence of shocks

We demonstrate our post-shock forecasting methodology on a time series of Conoco Phillips share prices after the occurrence of a structural shock. Conoco Phillips is a large oil and gas resources company (ConocoPhillips, 2020). The particular post-shock response that we predict happened after trading ended on Friday March 6th, 2020 and before trading began on Monday March 9th, 2020. It is reasonable that the timing of this shock is known, several events occurred over the trading weekend which had an impact on stock markets and the oil markets. For example, Russia and OPEC began a battle for global oil price control on Sunday, March 8th (Sukhankin, 2020), and several US states began declaring state of emergencies in response to the evolving coronavirus pandemic (New York State Government, 2020; Alonso, 2020).

We collect our data through the use of `quantmod` package (Ryan, Ulrich, Thielen & Teetor, 2017) in R. In this analysis we make the following design considerations:

(1) **Selection of model**. We will use an AR(1) model to forecast Conoco Phillips stock price. This model has been shown to beat no-change forecasts when predicting oil prices over time horizons of one and three months (Alquist, Kilian & Vigfusson, 2013). For these reasons, we will consider 30 pre-shock trading days and we will forecast the immediate shock effect. All

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

(2) **Selection of covariates**. We consider different covariates for the model of α_i and $y_{i,t}$. The model of α_i incorporates lagged daily S&P 500 index prices, West Texas Intermediate (WTI) crude oil prices, dollar index, 13 Week treasury bill rates, and Chicago Board Options Exchange volatility index (VIX). The model for $y_{i,t}$ disregards VIX. This is because VIX is a metric for capturing market risk and sentiment, which is highly influential for the shock effect model.

(3) **Construction of donor pool**. Our donor pool consists of Conoco Phillips shock effects observed in the past. We consider shock effects which occurred on March 14, 2008, several days in September, 2008, and November 27, 2014. The first sets of shock effects were observed during recessions that possessed similar characteristics to the current recession. In particular, all of these recessions were predicated by an inversion of the yield curve (Bauer & Mertens, 2018). These 2008 shock effects correspond to the collapse of Bear Stearns, the placement of Fannie May and Freddie Macbook in conservatorship on September 7th, the collapse of Lehman Brothers on September 15th, and the closing of Washington Mutual on September 25th (Shorter, 2008; Ewing & Malik, 2013; Dwyer & Tkac, 2009; Longstaff, 2010). The last shock effect corresponds to an OPEC induced supply side shock effect (Huppmann & Holz, 2015).

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. The estimated shock-effects for α_i are $-0.922, -7.063, -5.777, -6.395, -4.207$ for $i = 2, \ldots, 6$, respectively. Under $\mathcal{M}_2$, we computed $\hat{\alpha}_{\text{adj}}$, weighted adjustment $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$. Note that non-uniqueness problems will not occur in this analysis since the conditions of Proposition 1 are satisfied. To avoid the effect of unit differences on weighting, we centered and scaled the covariates in weights computation but
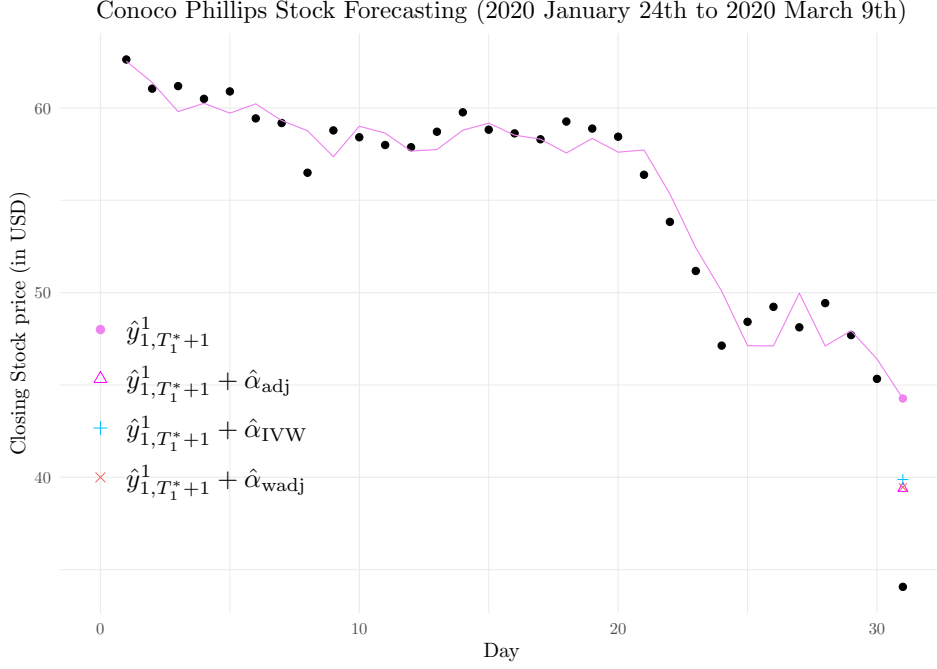
Figure 2: March 9th, 2020 post-shock forecasts for Conoco Phillips stock price. Black dots and magenta line represent real data and fitted line, respectively.

not in the model of $y_{i,t}$. For $\hat{\alpha}_{\text{wadj}}$, we observe that $\mathbf{W}^* = (0.000, 0.000, 0.000, 0.273, 0.727)$ and $\|\mathbf{X}_1 - \hat{\mathbf{X}}_1(\mathbf{W}^*)\|_2 = 3.440$. Note that the norm is computed using the k-dimensional Euclidean metric. The solution $\mathbf{W}^*$ suggests that the shock effect of interest is very similar to the September 25, 2008 shock effect and the November 27, 2014 shock effect.

The resulting shock effect estimates are $\hat{\alpha}_{\text{adj}} = -4.872$, $\hat{\alpha}_{\text{wadj}} = -4.805$, and $\hat{\alpha}_{\text{IVW}} = -4.384$. Using the bootstrap procedure $\mathcal{B}_f$, we estimated parameters for risk-reduction propositions and risk-reduction quantities proposed in Section 3. The estimated bootstrap variances for $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ are 0.419, 0.559, and 0.667, respectively. We verify the consistency of the result yielded by risk-reduction propositions with the reality as below.

We can see from Figure 2 that $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$ and $\hat{\alpha}_{\text{IVW}}$ perform decently well. Our estimation framework misses the yet to be observed pot-shock response, however they perform much better than unadjusted forecasts that do not account for shock effects. The unadjusted forecast misses the post-shock response by 9.870 dollars whereas the use of $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ misses by 5.324, 5.392, and 5.813 dollars, respectively.

The true shock effect is not fully recovered by $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$. This may be a result of

a poorly constructed donor pool. The shock on March 9th, 2020 is 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. In any event, the shock on March 9th, 2020 was the largest price shock to Conoco Phillips shares by a wide margin, even after adjusting for inflation.

From another perspective, it is possible that the stock of Conoco Phillips actually experienced multiple shocks on 2020 March 9th. For example, Kilian (2009) studied the effect that different supply and demand shocks have on oil prices through a vector autoregressive model. Their model postulates an additive nature of shock effects, although the additivity parameters requires estimation in their context. Motivated by Kilian (2009), we also studied additive shock effect estimators where the shock effects corresponding to separate supply and demand shocks are added to estimate the unknown shock effect. The supply shock donor pool consists of the November 27, 2014 shock effect; and the demand shock donor pool consists of the remaining shock effects. The additive adjustment estimator computed by adding the $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ estimators for the demand and supply shock effects only, respectively, miss the post-shock Conoco Philips share price by 0.951, 0.405, and 1.460 dollars. These additive adjustment estimators do extremely well in this additive shock setting.

5.2. Forecasting Apple stock in the presence of shocks

Apple Inc. is an American technology company that develops a range of electronics and software with typical products such as, iPhone, iPad, Mac, and Apple Watch. Every year, Apple Inc. releases a new series of MacBook computers to the public. Typically, after the release, the stock price of Apple Inc. is expected to experience a shock since Macbook sales is one of the most profitable line of Apple Inc.'s business and investors are likely to adjust their investment decisions based on the new features introduced. For example, the sales of Macbooks account for 9.8% of its revenue (Delventhal, 2020). In this example, we are interested in forecasting the stock price of Apple Inc. after the release of the latest MacBook on November, 10th, 2020. In recent years, we observe that it takes two or more days (Why is it two or more days and not just two days? It took two trading days for shock effects for 2017 and 2018 but 6 trading days for 2019. How can we be certain that we want the shock effect on November 11th instead of November 12th or some later date? We cannot know how long it took for the shock effect in 2020 to take an effect. From the past data (i.e., data in 2017 and 2018), I think 2 trading days are a safe guess.) for the shock

26

to take effect on the stock price. Therefore, we are interested in the stock price of Apple Inc. on the day after the release date, i.e., on November 11th, 2020. The detailed considerations of our analysis are outlined as below.

(1) **Selection of model.** We use an AR(1) model to forecast the stock price of Apple Inc. We consider 20 trading days data for the time series of interest and time series in the donor pool. All the estimates will be adjusted for inflation. Since we are interested in the stock price of the day after the release date, consider $D_{i,t} = I(t \in \{T_i^* + 1, T_i^* + 2\})$. As a result, we consider weighting with respect to

$$\mathbf{X}_1 = \mathbf{x}'_{1, T_1^*+1:T_1^*+2} \quad \text{and} \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}'_{2, T_2^*+1:T_2^*+2} \\ \vdots \\ \mathbf{x}'_{n+1, T_{n+1}^*+1:T_{n+1}^*+2} \end{pmatrix},$$

where $\mathbf{x}_{i, T_i^*+1:T_i^*+2} \in \mathbb{R}^{p \times 2}$ denotes the matrix joining both covariate column vectors corresponding to time points $t = T_i^* + 1$ and $t = T_i^* + 2$. You need to define the space of possible values for possible weighs in this new context, it is not immediately clear what you did. In this case, $\mathcal{W} = \{\mathbf{W} \in [0,1]^n : \mathbf{1}'_n \mathbf{W} = 1\}$ and we aim to find $\mathbf{W}^* \in \mathcal{W}$ with $\mathbf{W}^* = (w_2^*, \ldots, w_{n+1}^*)'$ such that

$$\mathbf{x}_{i, T_i^*+1:T_i^*+2} = \sum_{i=2}^{n+1} w_i^* \mathbf{x}_{i, T_i^*+1:T_i^*+2}.$$

and $\mathbf{W}^*$ is estimated by

$$\mathbf{W}^* = \underset{\mathbf{W} \in \mathcal{W}}{\arg\min} \left\| \mathbf{x}_{i, T_i^*+1:T_i^*+2} - \sum_{i=2}^{n+1} w_i^* \mathbf{x}_{i, T_i^*+1:T_i^*+2} \right\|_{2p}.$$

The model $\mathcal{M}_2$ is then changed to

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

where $\delta_i \in \mathbb{R}^{p \times 1}$ is not assumed to be random. The above does not make sense, $\delta_i' \mathbf{x}_{i, T_i^*+1:T_i^*+2}$ is not a scalar, and $y_{i,t}$ is not a vector, right? It seems that, as currently written, α_i should be replaced by $\alpha_{i,t}$, for $t = T^* + 1, T^* + 2$, where it is understood that α_{i,T^*+1} is close to zero. But I am not even sure if that is correct, why is the shock effect for the irrelevant

27

nuisance next day response even considered at all? What did you actually do here? I change the model accordingly where we cannot assume δ_i is random anymore in this context and it has to be particular δ_i such that the model holds. It is also possible that we can use $\alpha_{i,t}$. By taking it two days to take effect, I mean α_{i,T_i^*+1} is small but not close to zero and it should be taken into consideration together with α_{i,T_i^*+2}. I personally prefer α_i instead of $\alpha_{i,t}$ since α_i is more intuitive in that there is only one gross shock effect. Or, there is another possible approach. Recall that we can use different covariates in the model of α_i and in the model of $y_{i,t}$. So, we can only use $\mathbf{x}_{i,T_i^*+1}$ to construct the weights. Please let me know what you think and which model you prefer.

(2) **Selection of covariates.** We consider lagged S&P 500 index prices and 13 Week treasury bill rates as covariates. In this example, the weighting method will require future covariates in computation. As a result, forecasting of future covariates is needed. We used AIC to select an AR(3) model for forecasting the S&P 500 index prices and an AR(1) model for forecasting the 13 Week treasury bill rates. We used Ljung-Box test to test model adequacy for 10 lags and failed to reject the null hypothesis that autocorrelation does not exist in the residuals series of the corresponding models. We then predict the future covariates.

(3) **Construction of donor pool.** Our donor pool consists of the time series data of Apple Inc. with shock-effect date in the middle of 15 trading days for 2017, 2018, and 2019. The specific shock-effect dates are June 6th, 2017, October 29th, 2018, and June 3rd, 2019, respectively. We only include the last three Macbook releases in our donor pool because those releases are similar to the current release in the sense that Apple inc. has only updated the hardware of the computer without much redesign in these releases (Lee, 2017; Hanson, 2018, 2019; Newsroom, 2020).

The covariates and response of time series in the donor pool are adjusted for inflation. The estimated shock effects α_i are $0.864, 0.905, 1.129$ for $i = 2, 3, 4$, respectively. We centered and scaled the covariates in weights computation but not in the model of $y_{i,t}$. As a result, we observe $\mathbf{W}^* = (0.664, 0.000, 0.336)$ It is still a little confusing where this came from since it is not defined in this setting. and $\|\mathbf{X}_1 - \hat{\mathbf{X}}_1(\mathbf{W}^*)\|_2 = 3.674$. It appears that the shock effect of interest is similar to 2017 and 2019 shock effects. The resulting shock effect estimates are $\hat{\alpha}_{\text{adj}} = 1.071$, $\hat{\alpha}_{\text{wadj}} = 1.066$,
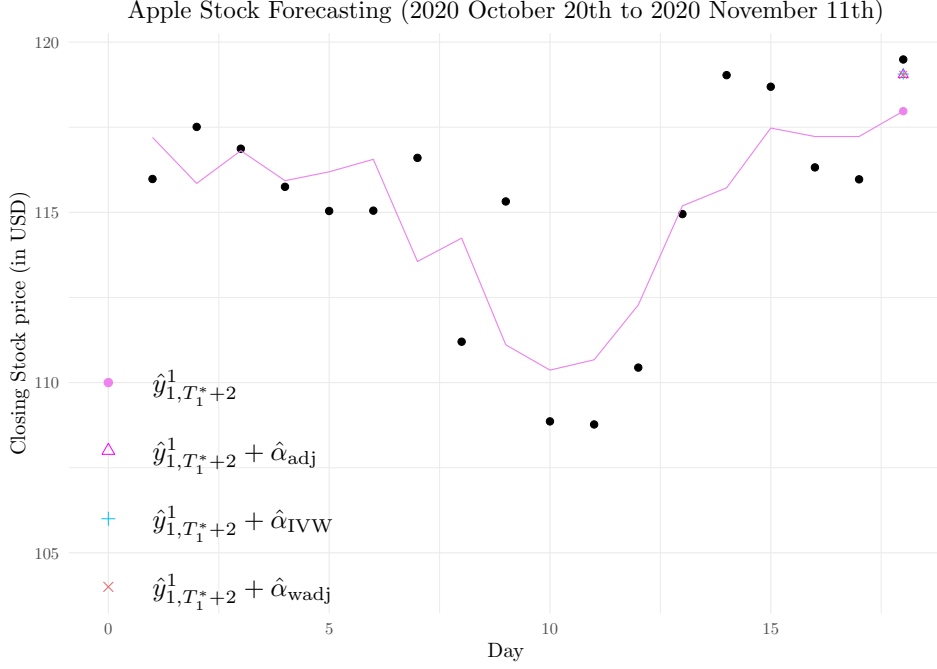
Figure 3: November 11th, 2020 post-shock forecasts for stock price of Apple Inc. Black dots and magenta line represent real data and fitted line, respectively.

and $\hat{\alpha}_{\text{IVW}} = 1.090$. The bootstrap estimates variances for variances for $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ are 0.192, 0.095, and 0.091, respectively. Our risk-reduction propositions yield that those shock-effect estimators reduce the post-shock forecast risk. As we can see from Figure 3, the post-shock forecasts perform well. The unadjusted forecast misses the post-shock by 1.518 dollars, while the use of $\hat{\alpha}_{\text{adj}}$, $\hat{\alpha}_{\text{wadj}}$, and $\hat{\alpha}_{\text{IVW}}$ only miss by 0.447, 0.427, and 0.452 dollars, respectively. Note that the 1-ahead forecast error for the covariates contribute to some of the forecast risk.

6. Discussion

We developed a methodology for forecasting post-shock response values after the occurrence of a structural shock. Our methodology is as follows: construct a synthetic panel of 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. There have been several other recent similar methods developed for forecasting COVID-19 cases. For example, Lee, Lei & Mallick (2020) constructed a Bayesian hierarchical model embracing data integration to improve predictive precision of COVID-19 infection trajectories for different coun-

tries. A similar setup may be appropriate for post-shock forecasting but may be too dependent upon model specification for the shock distribution. Plessen (2020) employed a data-mining approach to combine COVID-19 data from different countries as input to predict global net daily infections and deaths of COVID-19 using 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. Agarwal et al. (2020) proposed a model-free synthetic intervention method to predict unobserved potential outcomes after different interventions given a donor pool of observed outcomes with given interventions. They also provided 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 Agarwal et al. (2020) is quite general, there is no guarantee for theoretical properties in prediction without assuming any distributional structure.

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. Generally, multiple shock effects can be nested within a time series; and time series in the donor pool can be dependent. As an example, we considered a dependency structure for the September 2008 shock effects in our analysis of Conoco Phillips stock. But we note that consistency estimates from LOOCV 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 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 $y_{i,t}$. Additionally, α_i can be heteroskedastic; i.e., they can have different variances. In this scenario, all the theoretical properties still hold though variance expressions in Section 3.1 may not apply. 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. Although our work is developed for time-series or AR(p) models, in fact, it can be

generalized to any similar setting with a model of the response, whose parameters can be estimated unbiasedly, an additive shock-effect structure, and the structure that the time series in the donor pool are independent of the one of interest.

We can extend the proposed nowcasting to forecast the response for m periods, where $m \geq 2$. Without loss of generality, we assume $m = 2$. There are two cases. First, suppose that there is no additional shock after $T_1^* + 1$. Then, the forecast boils down to standard m-ahead forecast in time series. Second, assume that the time series under study will experience a shock at $T_1^* + 2$, say $\alpha_1^{(2)}$. Under our framework, users may want to construct another set of donor pool with time series that experience shock similar to $\alpha_1^{(2)}$. Let the resulting estimator using one of three weighting scheme be $\hat{\alpha}_1^{(2)}$. The forecast at $t = T_1^* + 2$ is

$$\hat{y}_{1,T_1^*+2}^2 = \hat{\eta}_1 + \hat{\phi}_1 \hat{y}_{1,T_1^*+1}^2 + \hat{\theta}_1' \mathbf{x}_{1,T_1^*+2} + \hat{\alpha}_1^{(2)},$$

where $\hat{y}_{1,T_1^*+1}^2$ is defined as in Section 2.2. These two cases can be generalized to forecasting $m \geq 2$ more periods. The scenarios for forecasting in AR(p), or other general models are analogous.

In the setup, we assume that at $t = T_1^*$, the user knows there will be a shock at $t = T_1^* + 1$. The methodology can be extended to more general timing scenarios. Suppose that we know at T_1^* that a shock is about to occur at $T_1^* + d$ for $d \geq 1$ and $d \in \mathbb{N}$, and that we have information about the future covariates for $t = T_1^* + 1, \ldots, T_1^* + d$. We can make a scalar adjustment for original forecast at $T_1^* + d$ using shock-effect estimator constructed from donor pool.

Our bootstrap procedures can be extended to approximate the distribution of shock effect estimators from more general time series. If the data are subject to heteroskedasticity of unknown form, bootstrapping tuples of regressands and regressors proposed by Freedman (1981) is robust in this situation with asymptotic validity in autoregressive models established by Gonçalves & Kilian (2004). If serial correlation exists in the data, various block bootstrapping procedures (Künsch, 1989; Liu, Singh et al., 1992) can be possible reasonable alternatives. Note that the pseudo time series generated by our proposed residual bootstrap are not stationary. If stationarity is of concern, one can be referred to the stationary bootstrap in Politis & Romano (1994) for stationary and weakly dependent time series. Nevertheless, it was shown that approximation accuracy might be a cost for the stationary bootstrap in autoregressive models in finite sample (Berkowitz, Birgean & Kilian, 2000). More work related to bootstrapping time series can be referred to Chapters 3 and 4 in Politis, Romano & Wolf (1999), Berkowitz & Kilian (2000), and Chapter 12 in Kilian

& Lütkepohl (2017). It is up to users in terms of selecting which procedure to choose but under different assumptions on the time series.

Construction of the donor pool is a critical step that directly matters in prediction. In our model, shocks from candidate time series in the donor pool, together with the shock in the time series of interest, should come from a family of distributions that allow for varying variances and varying means that are linear functions of common covariates. Remark 1 shows that common covariates are not restrictive with structural zero in coefficients. However, it is difficult to verify such assumption or structure in practice. In principle, one should select identities that come from a common population, which is also the one the identity under study come from, as candidates in donor pool. This selection criterion is often adopted in the literature of comparative case studies. For example, Card (1990) used a set of cities in southern U.S. to estimate the effect of 1980 Mariel Boatlift on the Miami labor market. Abadie & Gardeazabal (2003) utilized a donor pool of two Spanish regions to approximate the effect of terrorism on the economic growth of Basque Country. To estimate the effect of Proposition 99, a large tabacco control program in California in 1988, on the annual per-capita cigarette sales of 2000, Abadie et al. (2010) constructed a donor pool consisting of US states, which did not implement *any* large-scale tobacco control program during their sampling period. Common features of those studies are that they all carefully constructed the donor pool and discarded any suspicious outliers.

7. Supplementary material

Supplementary material includes the results of likelihood ratio test for independence among nested shock effects in Section 5.1, the algorithmic formulation of $\mathcal{B}_u$ and $\mathcal{B}_f$, simulation results for $\mathcal{M}_1$, simulation results for an example of the boundary case $p < n$, and R codes for corresponding simulations and data analysis.

8. Appendix

8.1. Justification of Expectation of $\hat{\alpha}_{adj}$ and $\hat{\alpha}_{wadj}$

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

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

$$\mathrm{E}(\hat{\alpha}_{\mathrm{adj}}) = \frac{1}{n} \sum_{i=2}^{n+1} \mathrm{E}(\mathrm{E}(\hat{\alpha}_i | \Theta)) = \mu_\alpha$$

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

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

$$\begin{aligned}
\mathrm{E}(\hat{\alpha}_{\mathrm{wadj}}) &= \mathrm{E}\left\{ \mathrm{E}\left(\sum_{i=2}^{n+1} w_i^* \hat{\alpha}_i | \Theta \right) \right\} \\
&= \mathrm{E}\left(\sum_{i=2}^{n+1} w_i^* \alpha_i \right) \\
&= \mathrm{E}\left\{ \sum_{i=2}^{n+1} w_i^* \left[\mu_\alpha + \delta_i' \mathbf{x}_{i,T_i^*+1} \right] \right\} \\
&= \mu_\alpha + \mu_\delta' \sum_{i=2}^{n+1} w_i^* \mathbf{x}_{i,T_i^*+1} & (\mathbf{W} \in \mathcal{W}) \\
&= \mu_\alpha + \mu_\delta' \mathbf{x}_{1,T_1^*+1}. & (\text{from (5)})
\end{aligned}$$

Similarly,

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

8.2. Justification of Variance of $\hat{\alpha}_{\mathrm{adj}}$ and $\hat{\alpha}_{\mathrm{wadj}}$

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

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

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

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

33

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

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

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

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

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

$$\text{Var}(\hat{\alpha}_{\text{adj}}) = \frac{\sigma^2}{n^2} \sum_{i=2}^{n+1} \text{E}\big\{(\mathbf{M}_i'\mathbf{M}_i)_{22}^{-1}\big\}$$
$$+ \frac{1}{n^2}(\mathbf{x}_{i,T_i^*+1}'\Sigma_\delta\mathbf{x}_{i,T_i^*+1} + \sigma_\alpha^2).$$

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

8.3. Proofs for lemmas and propositions

Proof of Proposition 1 The proof of Li (2019) in Appendix A.2 and A.3 adapts easily to Proposition 1. $\qquad\square$

Proof of Proposition 2 The proof for unbiasedness follows immediately from discussions related to expectation in Section 3. 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.

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

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

Lemma 1. *The forecast risk reduction is* $R_{T_1^*+1,1} - R_{T_1^*+1,2} = \text{E}(\alpha_1^2) - \text{E}(\hat{\alpha} - \alpha_1)^2$ *for all estimators of* α_1 *that are independent of* Θ_1 *(see Section 2.1).*

Proof of Lemma 1 Define

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

$$- (\eta_1 + \phi_1 y_{1,T_1^*} + \theta_1' \mathbf{x}_{1,T_1^*+1}),$$

where Θ_1 is as defined in (3). Notice that

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

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

It follows that

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

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

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

It follows that

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

which finishes the proof. $\qquad\square$

Proof of Proposition 3 The proofs are arranged into two separate parts as below.

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

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=2}^{n+1}\hat{\alpha}_i\right) = \frac{1}{n}\sum_{i=2}^{n+1}\mathrm{E}(\hat{\alpha}_i) = \frac{1}{n}\sum_{i=2}^{n+1}\mathrm{E}(\mathrm{E}(\hat{\alpha}_i \mid \Theta))$$

$$= \frac{1}{n}\sum_{i=2}^{n+1}\mathrm{E}(\alpha_i) = \mu_\alpha = \mathrm{E}(\alpha_1),$$

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

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

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

$$= \mu_\alpha^2 - \mathrm{Var}(\hat{\alpha}_{\mathrm{adj}})$$

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

Proof for statement (ii): By Proposition 2, the property that $\hat{\alpha}_{\mathrm{wadj}}$ is an unbiased estimator of μ_α holds for $\mathcal{M}_1$. The remainder of the proof follows a similar argument to the proof of statement (i). $\qquad\square$

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

8.4. Tables for Section 4

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

n	σ_α	Guess			LOOCV with k random draws			Distance to y_{1,T_1^*+1}			
		$\delta_{\hat{\alpha}_{\text{adj}}}$	$\delta_{\hat{\alpha}_{\text{wadj}}}$	$\delta_{\hat{\alpha}_{\text{IVW}}}$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{adj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{wadj}}})$	$\bar{\mathcal{C}}^{(k)}(\delta_{\hat{\alpha}_{\text{IVW}}})$	Original	$\hat{\alpha}_{\text{adj}}$	$\hat{\alpha}_{\text{wadj}}$	$\hat{\alpha}_{\text{IVW}}$
5	5	1 (0)	1 (0)	1 (0)	0.91 (0.03)	0.91 (0.02)	0.9 (0.03)	53.23 (4.1)	15.88 (2.1)	16.78 (2.37)	15.82 (2.07)
	10	0.97 (0.03)	1 (0)	0.97 (0.03)	0.89 (0.03)	0.9 (0.03)	0.89 (0.03)	53.01 (4.47)	17.83 (2.38)	19.56 (2.56)	17.61 (2.36)
	25	0.93 (0.05)	0.97 (0.03)	0.93 (0.05)	0.74 (0.04)	0.81 (0.04)	0.75 (0.04)	53.38 (5.92)	26.44 (3.8)	29.06 (4)	26.11 (3.75)
	50	0.83 (0.07)	0.83 (0.07)	0.8 (0.07)	0.59 (0.05)	0.64 (0.05)	0.59 (0.05)	61.68 (7.73)	46 (6.31)	47.3 (7.14)	45.25 (6.32)
	100	0.7 (0.09)	0.87 (0.06)	0.7 (0.09)	0.53 (0.05)	0.54 (0.05)	0.53 (0.06)	85.68 (12.95)	87.25 (11.86)	87.07 (13.63)	85.65 (12.02)
10	5	1 (0)	1 (0)	1 (0)	0.91 (0.03)	0.92 (0.02)	0.91 (0.03)	48.18 (4.59)	20.47 (2.71)	19.13 (2.97)	20.53 (2.73)
	10	1 (0)	1 (0)	1 (0)	0.87 (0.03)	0.89 (0.03)	0.87 (0.03)	48.93 (4.71)	21.27 (2.6)	19.24 (3.03)	21.31 (2.61)
	25	0.93 (0.05)	0.97 (0.03)	0.9 (0.06)	0.74 (0.03)	0.77 (0.03)	0.74 (0.03)	51.18 (5.7)	26.68 (2.78)	27.17 (3)	26.53 (2.77)
	50	0.8 (0.07)	0.8 (0.07)	0.8 (0.07)	0.57 (0.04)	0.61 (0.04)	0.57 (0.04)	57.82 (7.81)	40.51 (4.37)	46.85 (4.02)	40.19 (4.27)
	100	0.73 (0.08)	0.93 (0.05)	0.7 (0.09)	0.51 (0.04)	0.51 (0.04)	0.5 (0.04)	79.3 (12.44)	72.33 (9.12)	88.81 (8.46)	71.83 (8.85)
15	5	1 (0)	1 (0)	1 (0)	0.94 (0.02)	0.95 (0.02)	0.94 (0.02)	51.11 (3.05)	14.94 (2.36)	14.09 (2.37)	15.09 (2.35)
	10	1 (0)	1 (0)	1 (0)	0.92 (0.02)	0.91 (0.02)	0.91 (0.02)	52.34 (3.19)	15.64 (2.73)	15.29 (2.74)	15.89 (2.68)
	25	0.93 (0.05)	0.97 (0.03)	0.93 (0.05)	0.73 (0.04)	0.76 (0.04)	0.73 (0.04)	56.03 (5.2)	25.49 (3.85)	27.38 (3.87)	25.27 (3.84)
	50	0.8 (0.07)	0.83 (0.07)	0.8 (0.07)	0.56 (0.04)	0.6 (0.04)	0.57 (0.04)	71.37 (7.76)	47.25 (6.42)	52.06 (6.57)	46.41 (6.43)
	100	0.63 (0.09)	0.67 (0.09)	0.63 (0.09)	0.52 (0.04)	0.42 (0.04)	0.53 (0.04)	111.91 (13.83)	92.95 (12.34)	103.13 (12.74)	91.07 (12.37)
25	5	1 (0)	1 (0)	1 (0)	0.93 (0.02)	0.94 (0.02)	0.93 (0.02)	47.79 (2.93)	14.83 (1.72)	14.83 (2.04)	14.76 (1.72)
	10	1 (0)	1 (0)	1 (0)	0.89 (0.03)	0.91 (0.02)	0.89 (0.03)	47.93 (3.25)	16.55 (1.89)	17.55 (2.12)	16.53 (1.88)
	25	1 (0)	1 (0)	1 (0)	0.83 (0.03)	0.82 (0.03)	0.83 (0.03)	49.78 (5.01)	26.42 (3.38)	29.11 (3.4)	26.45 (3.35)
	50	0.97 (0.03)	1 (0)	0.93 (0.05)	0.64 (0.05)	0.63 (0.05)	0.64 (0.05)	62.64 (7.4)	48.84 (6.4)	52.67 (6.28)	48.8 (6.35)
	100	0.83 (0.07)	0.8 (0.07)	0.83 (0.07)	0.57 (0.05)	0.59 (0.05)	0.59 (0.05)	103.37 (12.23)	97.81 (12.52)	102.4 (12.49)	97.63 (12.45)

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

| σ | σ_α | Guess | | | LOOCV with k random draws | | | Distance to y_{1,T_1^*+1} | | | |
		$\delta_{\hat\alpha_{\text{adj}}}$	$\delta_{\hat\alpha_{\text{wadj}}}$	$\delta_{\hat\alpha_{\text{IVW}}}$	$\bar{C}^{(k)}(\delta_{\hat\alpha_{\text{adj}}})$	$\bar{C}^{(k)}(\delta_{\hat\alpha_{\text{wadj}}})$	$\bar{C}^{(k)}(\delta_{\hat\alpha_{\text{IVW}}})$	Original	$\hat\alpha_{\text{adj}}$	$\hat\alpha_{\text{wadj}}$	$\hat\alpha_{\text{IVW}}$
5	5	1 (0)	1 (0)	1 (0)	0.94 (0.02)	0.97 (0.01)	0.95 (0.02)	48.84 (3.4)	15.72 (1.93)	15.32 (1.94)	15.72 (1.93)
	10	1 (0)	1 (0)	1 (0)	0.92 (0.03)	0.94 (0.02)	0.92 (0.03)	49.54 (3.73)	17.07 (2.01)	16.08 (2.16)	17.12 (1.99)
	25	0.87 (0.06)	1 (0)	0.87 (0.06)	0.77 (0.02)	0.81 (0.03)	0.77 (0.03)	51.78 (5.22)	24.78 (2.57)	26.12 (2.49)	24.62 (2.54)
	50	0.8 (0.07)	0.83 (0.07)	0.8 (0.07)	0.59 (0.05)	0.61 (0.04)	0.59 (0.05)	58.62 (7.74)	40.09 (4.7)	47.09 (4.09)	39.85 (4.58)
	100	0.7 (0.09)	0.93 (0.05)	0.73 (0.08)	0.5 (0.04)	0.49 (0.04)	0.51 (0.04)	82.66 (12.19)	72.83 (9.75)	89.03 (9.15)	72.31 (9.5)
10	5	1 (0)	1 (0)	1 (0)	0.91 (0.03)	0.92 (0.02)	0.91 (0.03)	48.18 (4.59)	20.47 (2.71)	19.13 (2.97)	20.53 (2.73)
	10	1 (0)	1 (0)	1 (0)	0.87 (0.03)	0.89 (0.03)	0.87 (0.03)	48.93 (4.71)	21.27 (2.6)	19.24 (3.03)	21.31 (2.61)
	25	0.93 (0.05)	0.97 (0.03)	0.9 (0.06)	0.74 (0.03)	0.77 (0.03)	0.74 (0.03)	51.18 (5.7)	26.68 (2.78)	27.17 (3)	26.53 (2.77)
	50	0.8 (0.07)	0.8 (0.07)	0.8 (0.07)	0.57 (0.04)	0.61 (0.04)	0.57 (0.04)	57.82 (7.81)	40.51 (4.37)	46.85 (4.02)	40.19 (4.27)
	100	0.73 (0.08)	0.93 (0.05)	0.7 (0.09)	0.51 (0.04)	0.51 (0.04)	0.5 (0.04)	79.3 (12.44)	72.33 (9.12)	88.81 (8.46)	71.83 (8.85)
25	5	0.97 (0.03)	1 (0)	0.97 (0.03)	0.7 (0.04)	0.76 (0.03)	0.7 (0.04)	50.09 (8.27)	38.1 (5.75)	37.98 (5.5)	38.44 (5.76)
	10	0.97 (0.03)	1 (0)	0.97 (0.03)	0.69 (0.04)	0.74 (0.03)	0.69 (0.04)	50.82 (8.15)	37.63 (5.61)	36.33 (5.64)	37.99 (5.61)
	25	0.9 (0.06)	0.9 (0.06)	0.9 (0.06)	0.62 (0.04)	0.64 (0.04)	0.61 (0.04)	53.01 (8.22)	38.78 (5.29)	36.82 (5.76)	38.88 (5.31)
	50	0.8 (0.07)	0.8 (0.07)	0.8 (0.07)	0.53 (0.04)	0.53 (0.04)	0.53 (0.04)	58.9 (9.12)	46.77 (5.54)	50.61 (5.71)	46.6 (5.51)
	100	0.7 (0.09)	0.9 (0.06)	0.67 (0.09)	0.51 (0.03)	0.56 (0.04)	0.5 (0.03)	79.64 (12.21)	72.76 (8.79)	89.48 (7.98)	72.17 (8.59)
50	5	0.77 (0.08)	0.8 (0.07)	0.77 (0.08)	0.6 (0.05)	0.63 (0.04)	0.59 (0.04)	71.22 (13)	70.31 (10.4)	72.3 (9.26)	70.79 (10.45)
	10	0.77 (0.08)	0.77 (0.08)	0.77 (0.08)	0.6 (0.05)	0.63 (0.05)	0.6 (0.05)	70.85 (12.91)	69.43 (10.22)	70.65 (9.29)	69.94 (10.26)
	25	0.7 (0.09)	0.73 (0.08)	0.7 (0.09)	0.54 (0.05)	0.56 (0.05)	0.55 (0.05)	70.32 (12.81)	67.61 (9.86)	67 (9.58)	68.06 (9.89)
	50	0.67 (0.09)	0.7 (0.09)	0.67 (0.09)	0.51 (0.05)	0.51 (0.04)	0.51 (0.05)	74.01 (12.66)	68.69 (9.63)	67.9 (10.21)	68.91 (9.64)
	100	0.5 (0.09)	0.6 (0.09)	0.47 (0.09)	0.47 (0.05)	0.49 (0.04)	0.45 (0.05)	92.71 (13.06)	83.66 (10.79)	94.53 (11.2)	83.56 (10.63)
100	5	0.47 (0.09)	0.47 (0.09)	0.47 (0.09)	0.51 (0.06)	0.57 (0.05)	0.49 (0.06)	130.47 (22.59)	135.16 (19.73)	141.42 (16.98)	136.3 (19.72)
	10	0.47 (0.09)	0.47 (0.09)	0.47 (0.09)	0.51 (0.05)	0.53 (0.05)	0.51 (0.06)	129.49 (22.49)	134.09 (19.52)	139.69 (16.96)	135.26 (19.51)
	25	0.47 (0.09)	0.43 (0.09)	0.5 (0.09)	0.53 (0.06)	0.57 (0.05)	0.51 (0.06)	127.17 (22.22)	131.43 (18.97)	134.47 (17.22)	132.42 (18.99)
	50	0.5 (0.09)	0.43 (0.09)	0.5 (0.09)	0.48 (0.06)	0.56 (0.04)	0.48 (0.05)	125.72 (21.8)	129.27 (18.16)	129.59 (17.59)	130.27 (18.15)
	100	0.47 (0.09)	0.47 (0.09)	0.43 (0.09)	0.43 (0.06)	0.57 (0.04)	0.47 (0.06)	128.38 (21.86)	130.05 (18.05)	131.83 (19.08)	130.33 (18.06)

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

n	σ_α	Guess			LOOCV with k random draws			Distance to y_{1,T_1^*+1}			
		$\delta_{\hat\alpha_{\text{adj}}}$	$\delta_{\hat\alpha_{\text{wadj}}}$	$\delta_{\hat\alpha_{\text{IVW}}}$	$\bar{C}^{(k)}(\delta_{\hat\alpha_{\text{adj}}})$	$\bar{C}^{(k)}(\delta_{\hat\alpha_{\text{wadj}}})$	$\bar{C}^{(k)}(\delta_{\hat\alpha_{\text{IVW}}})$	Original	$\hat\alpha_{\text{adj}}$	$\hat\alpha_{\text{wadj}}$	$\hat\alpha_{\text{IVW}}$
5	5	1 (0)	1 (0)	1 (0)	0.89 (0.03)	0.92 (0.02)	0.89 (0.03)	48.52 (3.93)	15.74 (2.34)	15.76 (2.34)	15.16 (2.24)
	10	1 (0)	1 (0)	1 (0)	0.89 (0.02)	0.91 (0.02)	0.89 (0.02)	47.7 (4.35)	18.26 (2.37)	18.97 (2.42)	17.68 (2.28)
	25	0.97 (0.03)	1 (0)	0.93 (0.05)	0.79 (0.03)	0.81 (0.03)	0.77 (0.03)	46.95 (6.11)	27.35 (3.81)	30.88 (3.84)	26.58 (3.83)
	50	0.8 (0.07)	0.93 (0.05)	0.8 (0.07)	0.62 (0.03)	0.65 (0.03)	0.63 (0.03)	56.85 (8.64)	46.96 (7.02)	52.92 (7.45)	45.95 (7.17)
	100	0.73 (0.08)	1 (0)	0.8 (0.07)	0.53 (0.04)	0.53 (0.04)	0.55 (0.04)	99.22 (12.84)	93.4 (12.97)	103.82 (14.1)	91.95 (13.34)
10	5	1 (0)	1 (0)	1 (0)	0.86 (0.03)	0.88 (0.02)	0.86 (0.03)	50.59 (5.24)	29.19 (5.2)	31.4 (5.28)	29.29 (5.22)
	10	1 (0)	1 (0)	1 (0)	0.82 (0.03)	0.84 (0.03)	0.82 (0.03)	51.17 (5.49)	31.55 (5.33)	33.91 (5.56)	31.7 (5.35)
	25	0.93 (0.05)	1 (0)	0.93 (0.05)	0.72 (0.04)	0.75 (0.04)	0.71 (0.04)	53.53 (6.58)	40.05 (6.03)	43.5 (6.66)	40.43 (6)
	50	0.87 (0.06)	0.97 (0.03)	0.87 (0.06)	0.55 (0.04)	0.58 (0.05)	0.55 (0.05)	62.45 (8.25)	55.56 (8.11)	62.15 (9.19)	56.12 (8.04)
	100	0.77 (0.08)	0.97 (0.03)	0.73 (0.08)	0.49 (0.05)	0.44 (0.05)	0.46 (0.05)	85.72 (12.73)	89.5 (13.37)	103.25 (15.07)	89.92 (13.29)
15	5	1 (0)	1 (0)	1 (0)	0.95 (0.02)	0.92 (0.03)	0.95 (0.02)	52.1 (2.96)	14.04 (1.78)	13.36 (2.07)	14.11 (1.76)
	10	1 (0)	1 (0)	1 (0)	0.92 (0.02)	0.9 (0.03)	0.93 (0.02)	52.25 (3.3)	15.12 (1.93)	14.24 (2.29)	15.18 (1.9)
	25	0.93 (0.05)	1 (0)	0.9 (0.06)	0.8 (0.03)	0.8 (0.03)	0.8 (0.03)	52.71 (5.28)	22.98 (2.9)	22.6 (3.42)	22.95 (2.88)
	50	0.7 (0.09)	0.9 (0.06)	0.7 (0.09)	0.65 (0.03)	0.65 (0.03)	0.65 (0.04)	58.65 (8.48)	39.51 (5.65)	40.8 (6.3)	39.35 (5.63)
	100	0.6 (0.09)	0.87 (0.06)	0.6 (0.09)	0.47 (0.05)	0.45 (0.04)	0.45 (0.05)	88.76 (13.66)	75.93 (11.52)	81.89 (12.19)	75.94 (11.41)
25	5	1 (0)	1 (0)	1 (0)	0.94 (0.02)	0.95 (0.02)	0.94 (0.02)	50.55 (2.9)	12.13 (1.77)	14.22 (1.96)	12.09 (1.77)
	10	1 (0)	1 (0)	1 (0)	0.93 (0.02)	0.95 (0.02)	0.93 (0.02)	49.13 (3.31)	14.78 (1.85)	18.21 (2)	14.75 (1.85)
	25	1 (0)	1 (0)	1 (0)	0.83 (0.02)	0.85 (0.02)	0.83 (0.03)	47.38 (5.13)	26.85 (3.33)	32.95 (3.59)	26.81 (3.32)
	50	0.97 (0.03)	1 (0)	0.97 (0.03)	0.61 (0.04)	0.71 (0.04)	0.62 (0.04)	56.73 (7.93)	50.96 (6.63)	60.62 (7.21)	50.88 (6.59)
	100	0.8 (0.07)	0.93 (0.05)	0.8 (0.07)	0.49 (0.05)	0.51 (0.04)	0.49 (0.05)	93.79 (14.67)	102.05 (13.37)	116.45 (15.19)	101.55 (13.38)

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

σ	σ_α	Guess			LOOCV with k random draws			Distance to y_{1,T_1^*+1}			
		$\delta_{\hat\alpha_{\text{adj}}}$	$\delta_{\hat\alpha_{\text{wadj}}}$	$\delta_{\hat\alpha_{\text{IVW}}}$	$\bar{C}^{(k)}(\delta_{\hat\alpha_{\text{adj}}})$	$\bar{C}^{(k)}(\delta_{\hat\alpha_{\text{wadj}}})$	$\bar{C}^{(k)}(\delta_{\hat\alpha_{\text{IVW}}})$	Original	$\hat\alpha_{\text{adj}}$	$\hat\alpha_{\text{wadj}}$	$\hat\alpha_{\text{IVW}}$
5	5	1 (0)	1 (0)	1 (0)	0.94 (0.02)	0.95 (0.02)	0.94 (0.02)	50.04 (3.65)	21.75 (4.34)	22.68 (4.41)	21.83 (4.34)
	10	1 (0)	1 (0)	1 (0)	0.93 (0.02)	0.92 (0.02)	0.93 (0.02)	49.93 (4.11)	24.3 (4.44)	25.72 (4.58)	24.4 (4.45)
	25	0.97 (0.03)	1 (0)	0.97 (0.03)	0.73 (0.04)	0.8 (0.03)	0.73 (0.04)	51.29 (5.39)	32.8 (5.25)	35.39 (5.78)	33.11 (5.21)
	100	0.77 (0.08)	0.97 (0.03)	0.7 (0.09)	0.49 (0.04)	0.44 (0.05)	0.47 (0.04)	82.55 (11.61)	84.26 (12.7)	98.66 (13.96)	84.62 (12.59)
10	5	1 (0)	1 (0)	1 (0)	0.86 (0.03)	0.88 (0.02)	0.86 (0.03)	50.59 (5.24)	29.19 (5.2)	31.4 (5.28)	29.29 (5.22)
	10	1 (0)	1 (0)	1 (0)	0.82 (0.03)	0.84 (0.03)	0.82 (0.03)	51.17 (5.49)	31.55 (5.33)	33.91 (5.56)	31.7 (5.35)
	25	0.93 (0.05)	1 (0)	0.93 (0.05)	0.72 (0.04)	0.75 (0.04)	0.71 (0.04)	53.53 (6.58)	40.05 (6.03)	43.5 (6.66)	40.43 (6)
	50	0.87 (0.06)	0.97 (0.03)	0.87 (0.06)	0.55 (0.04)	0.58 (0.05)	0.55 (0.05)	62.45 (8.25)	55.56 (8.11)	62.15 (9.19)	56.12 (8.04)
	100	0.77 (0.08)	0.97 (0.03)	0.73 (0.08)	0.49 (0.05)	0.44 (0.05)	0.46 (0.05)	85.72 (12.73)	89.5 (13.37)	103.25 (15.07)	89.92 (13.29)
25	5	0.97 (0.03)	1 (0)	0.97 (0.03)	0.7 (0.03)	0.73 (0.03)	0.71 (0.04)	57.87 (8.76)	50.31 (7.58)	57.25 (7.75)	50.53 (7.64)
	10	0.97 (0.03)	1 (0)	0.97 (0.03)	0.68 (0.04)	0.69 (0.04)	0.69 (0.04)	58.41 (9.11)	51.62 (7.91)	58.61 (8.28)	51.85 (7.98)
	25	0.93 (0.05)	0.97 (0.03)	0.9 (0.06)	0.63 (0.04)	0.68 (0.04)	0.63 (0.04)	62.02 (10.02)	59.08 (8.53)	65.55 (9.64)	59.68 (8.53)
	50	0.87 (0.06)	0.9 (0.06)	0.87 (0.06)	0.54 (0.04)	0.59 (0.04)	0.52 (0.04)	71.01 (11.73)	73.52 (10.18)	81.81 (11.98)	74.51 (10.09)
	100	0.73 (0.08)	0.87 (0.06)	0.73 (0.08)	0.47 (0.05)	0.45 (0.04)	0.5 (0.05)	95.93 (15.59)	104.49 (15.12)	119.38 (17.62)	105.72 (14.97)
50	5	0.8 (0.07)	0.77 (0.08)	0.8 (0.07)	0.52 (0.04)	0.49 (0.05)	0.52 (0.04)	85.95 (14.57)	90.02 (13.45)	103.03 (13.91)	90.08 (13.64)
	10	0.8 (0.07)	0.73 (0.08)	0.8 (0.07)	0.55 (0.05)	0.5 (0.04)	0.53 (0.05)	86.44 (14.95)	90.89 (13.79)	104.62 (14.28)	91.03 (13.98)
	25	0.77 (0.08)	0.77 (0.08)	0.77 (0.08)	0.53 (0.04)	0.46 (0.04)	0.53 (0.04)	90.26 (15.82)	95.36 (14.69)	109.53 (15.69)	95.79 (14.83)
	50	0.77 (0.08)	0.8 (0.07)	0.77 (0.08)	0.48 (0.05)	0.45 (0.05)	0.45 (0.05)	99.52 (17.26)	106.52 (16.17)	120.99 (18.13)	107.55 (16.19)
	100	0.57 (0.09)	0.77 (0.08)	0.63 (0.09)	0.41 (0.04)	0.45 (0.03)	0.41 (0.04)	123.11 (20.61)	135.24 (19.74)	151.62 (23.42)	137.21 (19.55)
100	5	0.63 (0.09)	0.57 (0.09)	0.63 (0.09)	0.48 (0.05)	0.48 (0.03)	0.47 (0.05)	156.82 (26.36)	170.06 (25.92)	196.4 (26.55)	170.13 (26.27)
	10	0.63 (0.09)	0.57 (0.09)	0.67 (0.09)	0.46 (0.05)	0.47 (0.03)	0.47 (0.05)	157.3 (26.76)	170.93 (26.21)	197.96 (26.87)	171.07 (26.56)
	25	0.67 (0.09)	0.63 (0.09)	0.67 (0.09)	0.44 (0.04)	0.5 (0.03)	0.45 (0.04)	160.32 (27.73)	173.66 (27.22)	202.61 (28.01)	174.28 (27.48)
	50	0.67 (0.09)	0.67 (0.09)	0.67 (0.09)	0.39 (0.04)	0.43 (0.03)	0.38 (0.04)	166.98 (29.35)	182.83 (28.37)	210.38 (30.48)	183.67 (28.61)
	100	0.6 (0.09)	0.67 (0.09)	0.53 (0.09)	0.4 (0.04)	0.45 (0.04)	0.41 (0.05)	188.29 (32.07)	203.7 (31.63)	233.44 (35.31)	205.4 (31.72)

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, *105*, 493–5ARTICLE05.

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, *93*, 113–132.

Agarwal, A., Alomar, A., Sarker, A., Shah, D., Shen, D., & Yang, C. (2020). Two burning questions on covid-19: Did shutting down the economy help? can we (partially) reopen the economy without risking the second wave? *arXiv preprint arXiv:2005.00072*, .

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

Alquist, R., Kilian, L., & Vigfusson, R. J. (2013). Forecasting the price of oil. In *Handbook of economic forecasting* (pp. 427–507). Elsevier volume 2.

Andrews, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, *68*, 399–405.

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

Bauer, M. D., & Mertens, T. M. (2018). Economic forecasts with the yield curve. *Federal Reserve Bank of San Francisco Economic Letter*, (pp. 1–5).

Baumeister, C., & Kilian, L. (2014a). A general approach to recovering market expectations from futures prices with an application to crude oil, .

Baumeister, C., & Kilian, L. (2014b). Real-time analysis of oil price risks using forecast scenarios. *IMF Economic Review*, *62*, 119–145.

Berkowitz, J., Birgean, I., & Kilian, L. (2000). On the finite sample accuracy of nonparametric resampling algorithms for economic time series. *Advances in Econometrics*, (pp. 77–105).

Berkowitz, J., & Kilian, L. (2000). Recent developments in bootstrapping time series. *Econometric Reviews*, *19*, 1–48.

Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *Industrial and Labor Relations Review*, *44*, 245–257.

Castle, J. L., Clements, M. P., & Hendry, D. F. (2015). Robust approaches to forecasting. *International Journal of Forecasting*, *31*, 99–112.

Clements, M. P., Galvão, A. B. et al. (2019). *Measuring the Effects of Expectations Shocks*. Technical Report Economic Modelling and Forecasting Group.

Clements, M. P., & Hendry, D. F. (1996). Intercept corrections and structural change. *Journal of Applied Econometrics*, *11*, 475–494.

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

Croushore, D., & Evans, C. L. (2006). Data revisions and the identification of monetary policy shocks. *Journal of Monetary Economics*, *53*, 1135–1160.

Delventhal, S. (2020). Apple's 5 most profitable lines of business. https://www.investopedia.com/apple-s-5-most-profitable-lines-of-business-4684130#:~:text=For%20the%20fiscal%20year%202019,company's%20sales%3B%20the%20iPad%20accounted. Accessed on 2020-12-30.

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

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, (pp. 54–75).

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

Ewing, B. T., & Malik, F. (2013). Volatility transmission between gold and oil futures under structural breaks. *International Review of Economics & Finance*, *25*, 113–121.

Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, *9*, 1218–1228.

Gonçalves, S., & Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of econometrics*, *123*, 89–120.

Hanson, M. (2018). Macbook pro 2018 vs macbook pro 2017: is it a worthy upgrade? https://www.techradar.com/news/macbook-pro-2018-vs-macbook-pro-2017. Accessed on 2020-12-30.

Hanson, M. (2019). Macbook pro 2019 vs macbook pro 2018 – is it worth an upgrade? https://www.techradar.com/news/macbook-pro-2019-vs-macbook-pro-2018-is-it-worth-an-upgrade. Accessed on 2020-12-30.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

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

Huppmann, D., & Holz, F. (2015). *What about the OPEC Cartel?*. Technical Report DIW Roundup: Politik im Fokus.

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

Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge University Press.

Koop, G., & Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, *53*, 867–886.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, (pp. 1217–1241).

Lee, K. (2017). Macbook 2017 vs macbook 2016: what's changed? https://www.techradar.com/news/macbook-2017-vs-macbook-2016-whats-changed. Accessed on 2020-12-30.

Lee, S. Y., Lei, B., & Mallick, B. K. (2020). Estimation of covid-19 spread curves integrating global data and borrowing information. arXiv:2005.00662.

Lehmann, E. L., & Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media.

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

Liu, L., Moon, H. R., & Schorfheide, F. (2020). Forecasting with dynamic panel data models. *Econometrica*, *88*, 171–201.

Liu, R. Y., Singh, K. et al. (1992). Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the limits of bootstrap*, *225*, 248.

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

Marden, J. I. (2015). Multivariate statistics: Old school. *University of Illinois*, .

Monti, F. (2008). Forecast with judgment and models. *National Bank of Belgium Working Paper*, .

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

Newsroom, A. (2020). Introducing the next generation of mac. https://www.apple.com/newsroom/2020/11/introducing-the-next-generation-of-mac/. Accessed on 2020-12-30.

Pesaran, M. H., Shin, Y., & Smith, R. P. (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American statistical Association*, *94*, 621–634.

Plessen, M. G. (2020). Integrated time series summarization and prediction algorithm and its application to covid-19 data mining. arXiv:2005.00592.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical association*, *89*, 1303–1313.

Politis, D. N., Romano, J. P., & Wolf, M. (1999). *Subsampling*. Springer Science & Business Media.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with pims data. *Marketing Science*, *12*, 103–124.

Ryan, J. A., Ulrich, J. M., Thielen, W., & Teetor, P. (2017). quantmod: Quantitative financial modelling framework. r package version 0.4-12.

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

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

Svensson, L. E. (2005). *Monetary policy with judgment: Forecast targeting*. Technical Report National Bureau of Economic Research.

Tsay, R. S. (1986). Time series model specification in the presence of outliers. *Journal of the American Statistical Association*, *81*, 132–141.