

# Referee report checklist

Jilei Lin and Daniel J. Eck

Thank you for the opportunity to resubmit at *International Journal of Forecasting*. Thank you to the reviewers for comments that led to an improved version of this manuscript. We address all of the reviewers' comments below. An outline of each change made or a rebuttal is provided for all of the reviewers' comments.

## Editors

---

We have now received feedback on your manuscript from two reviewers as well as an Associate Editor. The reviewers' reports are attached to this message, while the AE's comments are included below this letter.

Both reviewers and the AE agree that you address a relevant and interesting topic, and that your paper has the potential to provide a valuable addition to the forecaster's toolkit. That being said, the current version of the manuscript does not meet the publication standards of the IJF. A number of important points are raised by the reviewers that are either unclear or unresolved, and these need to be addressed. In addition, the Associate Editor provides some useful comments and suggestions, also pointing out that quite a substantial revision is required.

In sum, I have decided to give you the opportunity to revise and resubmit your manuscript. Please do note that this is a "major revision". This means that, while we feel that the work presented in the manuscript might be of interest for the journal, re-submitting your manuscript does not necessarily mean it will be accepted, eventually.

---

*Response:* Thank you for the opportunity to submit a revision. We include a point by point response for the comments and concerns raised by the reviewers.

## Associate Editor

**Comment:** Firstly, this contribution seems to me to be related to two existing strands of literature, neither of which is cited. Describing the contribution relative to these literatures might help some readers better appreciate the achievements. These are the literatures on "ad hoc" model adjustments for breaks, see e.g., [Clements and Hendry \[1996\]](#) for an early contribution, and [Castle et al. \[2015\]](#) for a more recent example. The second is to do with outliers (innovation, additive and level shifts) in ARMA models, see e.g., [Tsay \[1986\]](#). Comparing/contrasting

the approach in the paper with these literatures may be illuminating.

*Response:* Thanks for pointing it out. The discussion of those papers is included in the introduction section of the revised manuscript.

**Comment 2:** Secondly, a single empirical example is presented. I would like to see the results for a range of series - this would generate confidence in the proposed approach.

*Response:* We include another example in which we forecast the stock price of Apple Inc. after the latest release of their MacBook product in 2020. This example is included in Section 5.2 of the revised manuscript. In this example, the donor pool includes three past time series of Apple inc.'s stock price surrounding previous releases of the MacBook. In this example, the shock is with precedent and our method is expected to work well. This is because Apple Inc. is a well-established company that did not revolutionize the MacBook product in their last three releases. Apple Inc. mostly focused on upgrading the hardware in the MacBook, this same is true for the latest release. Our adjusted forecast performs very well in this setting, as expected. This example is representative for a class of problems involving shocks that are very similar to those in the past. In contrast, the example of Conoco Phillips is without precedent. In the past, Conoco Phillips did not experience a financial crisis shock caused by a pandemic supplemented with an oil price shock. In the Conoco Phillips example, our method is better than the original forecast, although we do not fully recover the yet to be observed stock price. We hope that the addition of the Apple example helps to generate confidence in our approach.

## Reviewer 1: Cover letter for authors.pdf

Thank you very much for taking the time to provide a thoughtful review of our manuscript. We think that these comments will go a long way to improve the exposition of our manuscript.

**Comment 1:** How could the method be applied to forecast two or more period after a shock?

*Response:* Thanks for pointing it out. The method can be extended to forecast two or more periods after a shock. There are two situations.

First, suppose that there is no additional shock after  $T_1^* + 1$ . From equation (1) in the manuscript, the forecast at time point  $T_1^* + 2$  is

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

where  $\hat{y}_{1,T_1^*+1}^2 = \hat{y}_{1,T_1^*+1}^1 + \hat{\alpha}_{T_1^*+1}$ ,  $\hat{y}_{1,T_1^*+1}^1$  is the forecast without adjustment,  $\mathbf{x}_{1,T_1^*+2}$  is either known or forecasted, and  $\hat{\alpha}_{T_1^*+1}$  is the shock-effect estimator for the shock at  $T_1^* + 1$ . The forecast for  $d > 2$  more periods is similarly stated as

$$\hat{y}_{i,T_1^*+d} = \hat{\eta}_1 + \hat{\phi}_i \hat{y}_{1,T_1^*+d-1} + \hat{\theta}_i' \mathbf{x}_{1,T_1^*+d},$$

where  $\mathbf{x}_{1,T_1^*+d}$  is either known or forecasted.

Second, suppose that there is one more shock effect after  $T_1^* + 1$ , say,  $\alpha_{T_1^*+d^*}$  for some  $d^* > 1$ . Without loss of generality, assume  $d^* = 2$ . Under our framework, users may want to construct another set of donor pool with time series that experienced shock effects similar to  $\alpha_{T_1^*+2}$ . Next, apply our proposed method to compute  $\hat{\alpha}_{T_1^*+2}$ , the estimator for the shock effect at  $T_1^* + 2$ . The forecast at the point  $T_1^* + 2$  will be,

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

These two cases can be further combined to generalize forecasting for two or more periods. We added aspects of this response to the Discussion.

**Comment 2:** Page 3, line 30: Blundell and Bond [1998] model is a fixed effect dynamic panel and estimate the parameters using the panel analysis. It would be nice if you clarify in what sense your work is similar to them.

*Response:* The model of Blundell and Bond [1998] is

$$y_{it} = \alpha y_{i,t-1} + \beta_1' x_{it} + \beta_2' x_{i,t-1} + \eta_i + v_{it}, \quad i = 1, \dots, N, \quad t = 2, \dots, T.$$

where  $|\alpha| < 1$ . In contrast, our model is

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

where  $|\phi_i| < 1$ . Our model is similar to the one of Blundell and Bond [1998] in the sense that we have the fixed-effect decomposition of the error term  $\eta_i + \varepsilon_{i,t}$  and the autoregressive structure. However, our model allows for  $\eta_i$  as random effects, different time lengths for each time

series, different autoregressive parameters, and a shock component. To avoid confusion, we remove this reference in the revised manuscript.

**Comment 3:** Page 4, line 50: Your paper considers forecasting a time-series model. It is confusing for readers to read your sentence In this article, we consider a dynamic panel data model...".

*Response:* Sorry for the confusion. It is corrected in the revised version.

**Comment 4:** Figures on page 5: Figure labels, (a) and (b), are missing. What does the black line in the top graph represent? Is it the realized observations?

*Response:* Thanks for pointing it out. The labels (a) and (b) are now added in the revised manuscript. The black line stands for the realized observations whereas the violet line stands for the fitted values.

**Comment 5:** Page 7, line 52: Are these time series estimates or panel data estimates? If panel estimators are used, then OLS estimators in this model results in inconsistency of the estimators. In that case, how wouldn't this inconsistency affect the results in section 3 about the unbiasedness?

*Response:* Sorry for not making it clear. They are time series estimates or the ordinary least squares (OLS) estimates but not panel data estimates. In fact, our method can be generalized to any model whose parameter can be estimated in an unbiased fashion. In the second paragraph of Section 6, we wrote

"Our methodology is developed for autoregressive models, but it can be generalized to any setting where model parameters can be estimated unbiasedly, there is an additive shock-effect structure, and the time series in the donor pool are independent from the one of interest."

**Comment 6:** Page 8, line 46: It is not clear to me how you estimate  $\alpha_i$ 's by OLS. Could you give the expressions for  $\hat{\alpha}_i$ ? Also, how can you identify  $\alpha_i$  from  $\eta_i$ ? Isn't  $\hat{\alpha}_{\text{adj}}$  an estimator of  $E(\alpha_1)$ ?

*Response:* Sorry for not making it clear. The estimation procedures can be explained in detail as follows. Note that for  $i = 2, \dots, n+1$ , the design matrix for  $i$ th time series is  $(\mathbf{1}_{T_i}, \mathbf{D}_i, \mathbf{Y}_i^{(1)}, \mathbf{X}_i)$ , where  $\mathbf{X}_i$  is the covariates of  $i$ th time series,  $\mathbf{D}_i \in \mathbb{R}^{T_i}$  with  $D_{i,t} = I(t = T_i^* + 1)$ , and  $\mathbf{Y}_i^{(1)}$  is the lagged response  $(y_{i,1}, \dots, y_{i,T_i})'$ . We define  $\mathbf{M}_i = (\mathbf{1}_{T_i}, \mathbf{D}_i, \mathbf{Y}_i^{(1)}, \mathbf{X}_i)$ . Then the expression for  $\hat{\alpha}_i$  is

$$\hat{\alpha}_i = [(\mathbf{M}_i' \mathbf{M}_i)^{-1} \mathbf{M}_i' \mathbf{Y}_i]_2, \quad \text{for } i = 2, \dots, n+1,$$

where  $[\cdot]_2$  denotes the second element of a vector. That is,  $\hat{\alpha}_i$  is the second element of the OLS estimate vector. Our model is

$$y_{i,t} = \eta_i + \alpha_i D_{i,t} + \phi_i y_{i,t-1} + \theta_i' \mathbf{x}_{i,t} + \varepsilon_{i,t}, \quad i = 1, \dots, n+1.$$

the intercept is denoted as  $\eta_i$ . Thus, we cannot identify  $\alpha_i$  from  $\eta_i$ . As defined in (4) in the manuscript, through averaging,  $\hat{\alpha}_{\text{adj}}$  is an estimator of  $E(\alpha_1)$ .

**Comment 7:** Page 9, line 10: It is not clear to me what  $\mathbf{U}_i$ 's are. An example could help understanding that better.

*Response:* Thanks for pointing it out. This quantity is the model (design) matrix corresponding to the  $i$ th time series. We clarify this quantity the revised version, and we change the notation to be  $\mathbf{M}_i$ . Our previous response describes  $\mathbf{M}_i$  in more detail.

**Comment 8:** Page 9, line 15: Why the closed form expressions for  $E(\hat{\alpha}_{\text{IVW}})$  and  $\text{Var}(\hat{\alpha}_{\text{IVW}})$  are not provided?

*Response:* In page 8, we wrote

"The inverse-variance weighted estimator is defined as

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

where  $\hat{\alpha}_i$  is the OLS estimator of  $\alpha_i$ ,  $\hat{\sigma}_i$  is the residual standard error from OLS estimation, and  $\mathbf{M}_i$  is the design matrix for OLS with respect to time series for  $i = 2, \dots, n+1$ . Note that since  $\sigma$  is unknown, estimation is required and the numerator and denominator terms are dependent in general."

To be more specific,  $\hat{\sigma}_{i\alpha}$  depends on  $\hat{\alpha}_i$ . As a result, the numerator and denominator of  $\hat{\alpha}_{\text{IVW}}$  are dependent. Without making distributional assumptions, it is impossible to evaluate  $E(\hat{\alpha}_{\text{IVW}})$  and  $\text{Var}(\hat{\alpha}_{\text{IVW}})$  in a closed form.

**Comment 9:** Page 13, line 14:  $\mu_\alpha$  is unknown, and in practice we have to estimate it. Would the results hold if one replaces  $\mu_\alpha$  with its estimate?

*Response:* We go for a plug-in approach. If we can estimate  $\mu_\alpha$  well, the results should hold with high probability. Section 3.2 is dedicated to this estimation problem. Using plug-in approaches, we get a decision rule to determine whether the scalar adjustment will be beneficial with correctness probability  $p$ . Section 3.3 details the leave-one-out cross validation procedures, which are tailored to our setup, to estimate  $p$ . As a result, users can be informed about the credibility of our methods.

**Comment 10:** Page 14, line 7: Shouldnt  $E(\alpha_1)$  be  $\mu_\alpha$ ? Previously  $\mu_\alpha$  was used.

*Response:* In the line 7 at Page 14,  $E(\alpha_1)$  is taken under  $\mathcal{M}_{21}$  or  $\mathcal{M}_{22}$ . In contrast,  $\mu_\alpha$  is used under  $\mathcal{M}_1$ . In Section 2.1, we define the distribution of  $\alpha_i$  as

$$\begin{aligned} \alpha_i &\sim \mathcal{F}_\alpha \quad \text{with} \quad E_{\mathcal{F}_\alpha}(\alpha_i) = \mu_\alpha, \text{Var}_{\mathcal{F}_\alpha}(\alpha_i) = \sigma_\alpha^2, \quad \text{iid under } \mathcal{M}_1 \\ \alpha_i &= \mu_\alpha + \delta_i' \mathbf{x}_{i, T_i^*+1} + \tilde{\epsilon}_i \quad \text{iid under } \mathcal{M}_{21} \text{ or } \mathcal{M}_{22}. \end{aligned}$$

In other words,  $\mathcal{M}_{21}$  or  $\mathcal{M}_{22}$  is an extended version of  $\mathcal{M}_1$  by allowing it to depend on some covariates. However, in general,  $E(\alpha_1) \neq \mu_\alpha$  under  $\mathcal{M}_{21}$  or  $\mathcal{M}_{22}$  due to the presence of  $\delta'_i \mathbf{x}_{i,T_i^*+1}$ .

**Comment 11:** Page 18, numerical example: The model setup does not contain the individual effects (intercepts) similar to the models in equations (1)-(2) on page 6. How would the result change if you include them with different variances?

*Response:* The result wouldn't change. In fact, in our proof listed in the Appendix, we do not make use of the assumption that  $\eta_i$  follows the same distribution. In other words, the results still hold when the individual effects have different variances.

**Comment 12:** Page 23, construction of donor pool: Your model assumes that the shocks have the same distribution. Given that these shocks are from different sources and happened because of various reasons, how could one justify them? Practically, there are many previous shocks. How can someone choose from them?

Our methodology requires that a practitioner can choose a suitable donor pool of candidate shock-effects. If this cannot be done, then the methodology may fail. In real-life applications, practitioners can use qualitative and quantitative information to inform their decisions about which time series are included in the donor pool. Consider our Conoco Phillips example. On March 9th 2020, Conoco Phillips was expected to experience a negative shock in the midst of the COVID-19 pandemic and an oil supply shock. While the COVID-19 pandemic is unprecedented, several, although not all, economic measurements surrounding March 9th 2020 resembled those in the days of the 2008 financial crises. On the other hand, the oil supply shock had precedent. As a result, we selected the shocks in the past corresponding to the 2008 financial crisis and a past 2014 oil supply shock.

**Comment 13:** Figure 2 on Page 24: Are the black dots the real data? What does the line represent?

*Response:* Thanks for pointing it out. We clarify this in the revised version. The black dots represent the real data. The line in violet color represents the fitted line of our model.

## Reviewer 2: MinimizingPostShockForestError\_IJF\_pdf

Thank you for your appreciation of the potential of our methods, and for your valuable detailed comments. We think these comments will help a lot in improving our manuscript.

**Comment 1:** The method hinges on the assumption that the time series forming the donor pool are independent from the time series under study prior to the shock. This seems to be a very restrictive assumption. The shocks that the donor pool have undergone in the past must be representative of the shock that the time series of interest is experiencing, but the series are independent. Given this restrictive assumption what practical applications are ruled out?

*Response:* Yes. It could be a restrictive assumption in a number of situations. Panel data applications are ruled out. Additionally, if time series in a donor pool are observed in almost concurrent time as the one under study (or, the time differences are not large enough), the independence assumption may not hold. However, if the time differences between the time series under study and those forming the donor pool are large enough, then the dependence should be negligible. This is the case in our Conoco Phillips and Apple examples. We include this discussion in the Introduction. Thank you for this comment.

**Comment 2:** It would be helpful to the reader to be very explicit about the timing of when information is realized. Is it correct that the researcher knows at  $T_1^*$  that there will be a shock at  $T_1^* + 1$ , and they know the type of shock that it will be in order to identify the relevant donor pool shocks, but they do not know the magnitude? This is quite an informational advantage over unanticipated shocks. In many situations one could imagine that the shock was unanticipated. Therefore, at  $T_1^* + 1$  there is a large forecast error (your model 1) but now information is available about the type of shock, so the proposed methodology of the shock effect estimator is applied at  $T_1^* + 2$ , yielding gains over the unadjusted forecast. The timing of information is crucial, but the methodology proposed in the paper seems to be applicable in many different timing scenarios.

*Response:* It is correct that the researcher knows at  $T_1^*$  that there will be a shock at  $T_1^* + 1$ . In practice, it usually happens to the case that when we realize that the shock is about to occur, we know the *general type* or *general cause* of the shock. The magnitude is not known. We admit that it is quite an informational advantage over unanticipated shocks. However, in practice, if a shock is unanticipated, we may not even be able to construct such a post-shock analysis since we do not know it is about to happen.

Our proposed methodology is applied at  $T_1^* + 1$ . You are right about the timing. Our methodology can be applicable in many different timing scenarios. Generally, 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  $t = T_1^* + 1, \dots, T_1^* + d$ . It is possible to generalize our method to this situation. Our new Apple example in Section 5.2 of the revised manuscript provides an example for the case  $d = 2$  with predicted covariates.

Thank you for your remark on timing scenarios, it led to an improved manuscript.

**Comment 3:** Related to the point above, the model applied is an AR(1) with contemporaneous regressors. This structure requires the covariates to be known in the future for *ex ante* forecasts, or for them to be forecast. The set-up assumes  $\theta_1$  to be well behaved and the covariates are not subject to shocks if they need to be forecast. Empirically, we could also think of cases where the mean shift  $D_{i,t}$  interacted with the covariates which would give an additional forecast error via  $\mathbf{x}_{1,T_1^*+1}$ . The theoretical results seem to hinge on known  $\mathbf{x}_{1,t}$  entering contemporaneously to derive unbiasedness. What are the implications of the covariates needing to be forecast?

*Response:* It is a very good point. Thanks for pointing it out. Indeed, our method assumes  $\mathbf{x}_{1,T_1^*+1}$  to be given. If it is to be forecasted with errors, it will possibly result in a biased estimation of  $E(\alpha_1)$  with  $\hat{\alpha}_{\text{wadj}}$  such that the evaluation of the proposition may not be reliable.

In fact, the situation when covariates are forecasted with errors is similar to the one when the variance of the shock effect is large. In both situations, the signal from covariates is weak. Our simulation discusses this situation by increasing  $\sigma_\alpha$ . It shows that a slightly weak signal will not compromise the correctness of our evaluated propositions a lot. However, if the forecast for the covariates is seriously bad (i.e., the signal from covariates is really weak), our result shows that the risk-reduction evaluation may be only as good as a random guess.

In the event that the mean shift  $D_{i,t}$  interacts with the covariates, one may be able to consider forecast scenarios (or conditional forecasts) [Baumeister and Kilian, 2014] to make post-shock predictions. In this setting, one would need to produce a grid of likely  $\mathbf{x}_{1,T_1^*+1}$  values to proceed. A post-shock prediction can be made for every hypothetical  $\mathbf{x}_{1,T_1^*+1}$  value forming the previously mentioned grid. This yields a range of post-shock predictions whose validity is subject to the accuracy of the  $\mathbf{x}_{1,T_1^*+1}$  values forming the grid. This may give users a sense of the variability of post-shock predictions due to forecast error from covariates affected from the shock.

**Comment 4:** Given the dynamic structure, what happens after  $T_1^* + 1$ ? Couldn't you use the donor pool shocks to estimate the shock transition (e.g. apply the shock effect estimator to the  $m$  periods following the shock), and produce an intercept correction dummy for the next  $m$  periods? It would be interesting empirically to see if the shock is anticipated to be permanent or transitory. I wonder if the theoretical statements of when forecast risk is minimized may be adapted to assess when transitory shocks die out such that the unadjusted forecast is preferable again?

*Response:* In our setup, we only model the time series under study up to  $t = 1, \dots, T_1^* + 1$ . We do not know what happened after  $T_1^* + 1$ . It is possible that there are no shocks or additional shocks after  $T_1^* + 1$ .

It is possible to use the donor pool to estimate the shock transition, and produce an intercept dummy for the next  $m$  periods. Assume there are no additional shocks after  $T_1^* + 1$ , and  $T_i^* + m \leq T_i$  for  $i = 2, \dots, n + 1$ . Consider the model

$$y_{i,t} = \eta_i + \theta_i' \mathbf{x}_{i,t} + \phi_i y_{i,t-1} + \beta_{i,T_i^*+d} I(t = T_i^* + d) + \varepsilon_{i,t},$$

for  $i = 2, \dots, n + 1$  and  $d = 1, \dots, m$ . Suppose we have information of the covariates up



to  $T_1^* + d$  either by given information or by forecasting. We can use our proposed synthetic control method to compute weights  $\mathbf{W}^*$  and OLS to estimate  $\beta_{i,T_i^*+d}$  to compute  $\hat{\beta}_{1,T_1^*+d}$ . As a result, we obtain  $\hat{\beta}_1^{(m)} = (\hat{\beta}_{1,T_1^*+d})_{d=1}^m$ . However, it cannot be a simple intercept dummy but instead an iterative adjustment in computing  $m$ -ahead forecasts. For example,

$$\hat{y}_{1,T_1^*+2}^{(1)} = \hat{y}_{1,T_1^*+2}^{(0)} + \hat{\beta}_{1,T_1^*+2},$$

where  $\hat{y}_{1,T_1^*+2}^{(0)}$  is the standard 2-ahead forecasts. One then may be able to use  $\hat{\beta}_1^{(d)}$  to judge whether the shock is transitory or permanent. We speculate that the transitory shock effect can be theoretically justified using similar arguments as those that appear in our manuscript, but we view this extension as outside the scope of our current paper. The SCM based approach of Agarwal et al. [2020] is an alternative that may also be able to assess whether or not a shock is transitory, although this adjustment is without any theoretical guarantees.

**Comment 5:** While the simulations were useful, I found the discussion in section 4 quite impenetrable. There are so many decisions that a researcher would need to make, including which weighting method and all the bootstrapping and leave-one-out decisions. It would be very helpful to the reader if you could give a summary of all your results stating your preferred user decisions for generic specifications.

*Response:* Thank you for pointing it out. We polished Section 4 for better readability in the revised manuscript. We summarize our results including preferred user decisions as below. There are mainly three decisions to make for an user: to choose

- among  $\hat{\alpha}_{\text{adj}}$ ,  $\hat{\alpha}_{\text{IVW}}$ , and  $\hat{\alpha}_{\text{wadj}}$
- between  $\mathcal{B}_f$  and  $\mathcal{B}_u$  (see detailed definition in Section 3.2)
- The  $k$  in the  $k$  random draws in LOOCV

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. These recommendations are included in Remark 3 at the end of Section 2.

The selection between  $\mathcal{B}_f$  and  $\mathcal{B}_u$  is more of a philosophical question. In Section 3.2, we noted

$\mathcal{B}_u$  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.

In Section 4.3, we found that  $\mathcal{B}_u$  is better than  $\mathcal{B}_f$  when the donor pool size is moderately small ( $n < 10$ ) and the signal from the covariates is large. Otherwise, they are nearly the same.

At the end of Section 4.3, we noted that when  $p < n$ , there are infinitely many solutions to  $\mathbf{W}^*$  such that  $\mathbf{W}^*$  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].  $\mathcal{B}_f$  is immune to this issue. However, our simulation results in the Supplementary Materials show that the non-uniqueness does not seriously compromise the inference.

In summary,  $\mathcal{B}_u$  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 due to theoretical reasons,  $\mathcal{B}_f$  is preferred. In other cases, the choice between  $\mathcal{B}_u$  and  $\mathcal{B}_f$  would be up to personal preferences related to philosophical differences mentioned above. This explanation is added to the end of Section 4.

Ideally,  $k = n$  is preferred. However, if the data size is large such that  $k = n$  is computationally expensive, we found that  $k = 5$  or  $k = 10$  works well. In practice,  $(k + 1)B$  number of bootstrap replications are implemented to complete the evaluation, where  $kB$  is from LOOCV to estimate the correctness parameter  $p$  and  $B$  is from ordinary bootstrap using the whole donor pool.

**Comment 6:** Section 2 is hard to read without defining the terms that come in the second paragraph of section 2.1 (lines 19-23 of p.6). I'd suggest moving this to the beginning of section 2.

*Response:* Sorry for not making it clear. The terms not defined in the caption of Figure 1 are now defined in the revised manuscript. For other parts, we find there are no terms that are not defined. The beginning parts of Section 2 are intended to give a general picture to the reader that our method is in fact general whereas the setup we investigate in Section 2.1 is a special case.

## References

- Anish Agarwal, Abdullah Alomar, Arnab Sarker, Devavrat Shah, Dennis Shen, and Cindy Yang. Two burning questions on covid-19: Did shutting down the economy help? can we (partially) reopen the economy without risking the second wave? *arXiv preprint arXiv:2005.00072*, 2020.
- Donald WK Andrews. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68(2):399–405, 2000.
- Christiane Baumeister and Lutz Kilian. Real-time analysis of oil price risks using forecast scenarios. *IMF Economic Review*, 62(1):119–145, 2014.
- Richard Blundell and Stephen Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143, 1998.
- Jennifer L Castle, Michael P Clements, and David F Hendry. Robust approaches to forecasting. *International Journal of Forecasting*, 31(1):99–112, 2015.

- Michael P Clements and David F Hendry. Intercept corrections and structural change. *Journal of Applied Econometrics*, 11(5):475–494, 1996.
- Ruey S Tsay. Time series model specification in the presence of outliers. *Journal of the American Statistical Association*, 81(393):132–141, 1986.