Response to Reviewer 2:
`MinimizingPostShockForestError_IJF_pdf`

Jilei Lin and Daniel J. Eck

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 $t = T_1^* + 1, \ldots, T_1^* + d$. 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 θ_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: Good point, thank you. 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. Moreover, the forecast error from covariates contributes to more error in the unadjusted forecast, and thus in the post-shock one.

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 σ_α. It shows that a slightly weak signal will not strongly compromise the correctness of our prospective methods. 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 gives users a sense of the variability of post-shock predictions due to forecast error from covariates that are also affected from the shock. We included this response to the Discussion, thank you again.

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 $t = 1, \ldots, T_1^* + 1$. 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 $i = 2, \ldots, n + 1$. 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 $i = 2, \ldots, n + 1$ and $d = 1, \ldots, m$. 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 β_{i,T_i^*+d} to compute $\hat{\beta}_{1,T_1^*+d}$. As a result, we obtain $\hat{\boldsymbol{\beta}}_1^{(m)} = (\hat{\beta}_{1,T_1^*+d})_{d=1}^m$. 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 $\hat{\boldsymbol{\beta}}_1^{(d)}$ 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 $\hat{\alpha}_{\mathrm{wadj}}$ [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.