Minimizing post shock forecasting error using disparate information

Daniel J. Eck

February 26, 2020

Abstract

We develop a forecasting methodology for time series data that is thought to have undergone a shock which has origins that have not been previously observed. We still can provide credible forecasts for a time series in the presence of such systematic shocks by drawing from disparate time series that have undergone similar shocks for which post-shock outcome data is recorded. These disparate time series are assumed to have mechanistic similarities to the time series under study but are otherwise independent (Granger noncausal). The inferential goal of our forecasting methodology is to supplement observed time series data with post-shock data from the disparate time series in order to minimize average forecast risk.

Things to do:

- get regression model interpretation of $\text{Var}(\hat{\alpha}_j)$.

- simulations for simple model with normal errors; model with normal errors and $p > 1$; model where shock effect is Gamma.

- more general panel model where covariates may be different.

- get references on time series pooling.

- investigate orthogonal parameterizations of regression model matrices in order to isolate estimation of $\text{Var}(\hat{\alpha}_j)$, so that the asymptotic estimate of $\text{Var}(\hat{\alpha}_j)$ is independent of other terms.

- expand method to weighted cases when the shock effect depends on the measured covariates.

- get a real dataset. Air conditiong units in hot humid summer vs salt sales in very icy/snoy winters and portable generators for hurricane/fires/tornados

1 Introduction

The technique of combining forecasts to lower forecast error has a rich history [Bates and Granger, 1969, Mundlak, 1978, Timmermann, 2006, Granger and Newbold, 2014]. The Introduction of Timmermann [2006] provided several reasons for combining forecasts. In particular, combining forecasts may be beneficial when: 1) the information set underlying individual forecasts is often unobserved to the forecast user; 2) different individual forecasts may be very differently affected by non-stationarities and model misspecifications; 3) different individual forecasts may be motivated by different loss functions [Timmermann, 2006, and references therein]. The setting for the forecast combination problem is that there are competing forecasts for a single time series. In this setting, one may desire combining forecasts as a method for lowering overall forecast error.

In this article we propse a new setting for the forecast combination problem. We will suppose that a time series of interest has recently undergone a structural shock that is not similar to anything observed in its past, and we desire reliable post-shock forecasts in this setting. It is unlikely that any forecast that previously gave successful predictions for the time series of interest will be able to accomondate the recent structural shock. However, it may be the case that disparate time series have previously undergone similar structural shocks. When this is so, one may be able to aggreagate the post-shock information from these disparate time series to aid the post-shock forecast for the time series under investigation.

2 Setting

We will suppose that an analyst has time series data $(y_{i,t}, \mathbf{x}_{i,t})$, $t = 1, \ldots, T_i$, $i = 1, \ldots, n+1$, where $y_{i,t}$ is a scalar response and $\mathbf{x}_{i,t}$ is a vector of covariates that revealed 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. To gauge the performance of a procedure that produces forecasts $\{\hat{y}_{1,t}, t = 1, 2, \ldots\}$ given time horizon T_1, we consider the average forecast risk

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

in our analyses. In this article, we consider a similar dynamic panel data model with autoregressive structure to that in Blundell and Bond [1998]. Our dynamic panel model includes an additional shock effect whose presence or absence is given by the binary variable $D_{i,t}$. Our dynamic panel model is

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

$t = 1, \ldots, T_i$ and $i = 1, \ldots, n+1$, where $D_{i,t} = 1(t > T_i^*)$, $T_i^* < T_i$ and $\mathbf{x}_{i,t} \in \mathbb{R}^p$, $p \geq 1$. We will consider the following random effects structure:

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

Throughout the rest of this article we show that the collection of disparate time series $\{y_{i,t}, t = 2, \ldots, T_i, i = 1, \ldots, n\}$ has the potential to improve the forecasts for $y_{1,t}$ when $t > T_1^*$ and no observations arre made for these time periods.

3 Adjustment via disparate infromation in dynamic panel model

Our first goal is to evaluate the forecast risk of competing forecasts which offer predicts for y_{1,T_1^*+1}, the response observed after the first post shock time period for out ime series of interest. The difficulty in this forecast stems from not observing α_1 and not having any readily infromation available to estimate it directly from model (1). We show that there is a balance between μ_α and σ_α^2 that allows us to lower the forecast risk for y_{1,T_1^*+1} by incorporating what is known about the other α_is when such information is available.

Conditional on all regression parameters, previous responses, and covariates, the response variable $y_{i,t}$ in model (1) has distribution

$$y_{i,t} \sim N(\eta_i + \alpha_i D_{i,t} + \phi_i y_{i,t-1} + \theta_i' \mathbf{x}_{i,t} + \beta_i' \mathbf{x}_{i,t-1}, \sigma^2).$$

All parameters in this model will be estimated with ordinary least squares (OLS) using historical data. In particular, let $\hat{\alpha}_i$, $i = 2, \ldots, n+1$ be the OLS estimate of α_i and define the adjusted α plugin estimator for time series $i = 1$ by,

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

where the $\hat{\alpha}_i$s in (2) are OLS estimators of all of the α_is. We can use $\hat{\alpha}_{\text{adj}}$ as an estimator for the unknown α_1 term for which no meaningful estimation information otherwise exists.

It is important to note that the adjustment $\hat{\alpha}_{\text{adj}}$ is a consistent estimator of μ_α, it is not an unbiased estimator for α_1, nor does it converge to α_1. Despite these inferential shortcomings, adjustment of the forecast for y_{1,T_1^*+1} through the addition of $\hat{\alpha}_{\text{adj}}$ has the potential to lower forecast risk in settings where μ_α is large relative to σ_α.

We will consider the candidate forecasts:

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

where $\hat{\eta}_1$, $\hat{\phi}_1$, $\hat{\theta}_1$, and $\hat{\beta}_1$ are all OLS estimators of η_1, ϕ_1, θ_1, and β_1 respectively. The first forecast ignores the information about the distribution of α_1 while the second forecast incorporates an estimate of μ_α that is obtained from the other individual forecasts under study. Note that the two forecasts do not differ in their predictions for $y_{1,t}$, $t = 1, \ldots T_1^*$. They only differ in predicting y_{1,T_1^*+1}. We want to determine when either $\hat{y}^1_{1,T_1^*+1}$ or $\hat{y}^2_{1,T_1^*+1}$ minimizes the forecast risk for $\hat{y}_{1,T_1^*+1}$. Let $R_{T_1^*+1,k} = \mathrm{E}(\hat{y}^k_{1,T_1^*+1} - y^k_{1,T_1^*+1})^2$, where $k = 1, 2$ and define the parameter sets

$$\begin{aligned}
\mathcal{H} &= \{(\eta_i, \phi_i, \theta_i, \beta_i, \alpha_i, \mathbf{x}_{i,t}, y_{i,t}); t = 1, \ldots, T_i, i = 2, \ldots, n+1\}, \\
\mathcal{H}_1 &= \{(\eta_1, \phi_1, \theta_1, \beta_1, \alpha_1, \mathbf{x}_{1,T_1^*+1}, \mathbf{x}_{1,t}, y_{1,t}); t = 1, \ldots, T_1^*\}.
\end{aligned} \tag{3}$$

Proposition 1 states when the incorporation of disparate information improves forecasting.

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

Proof. The forecast risk for y^2_{1,T_1^*+1} is

$$\begin{aligned}
R_{T_1^*+1,2} &= \mathrm{E}(\hat{y}^2_{1,T_1^*+1} - y^2_{1,T_1^*+1})^2 \\
&= \mathrm{E}\Big\{ \hat{\eta}_1 + \hat{\phi}_1 y_{1,T_1^*} + \hat{\theta}_1' \mathbf{x}_{1,T_1^*+1} + \hat{\beta}_1' \mathbf{x}_{1,T_1^*} + \hat{\alpha}_{\text{adj}}
\end{aligned}$$

$$
\begin{aligned}
&\quad -(\eta_1 + \phi_1 y_{1,T_1^*} + \theta_1' \mathbf{x}_{1,T_1^*+1} + \beta_1' \mathbf{x}_{1,T_1^*} + \alpha_1 + \varepsilon_{1,T_1^*+1})\big\}^2 \\
&= \mathrm{E}\Big\{(\hat{\eta}_1 + \hat{\phi}_1 y_{1,T_1^*} + \hat{\theta}_1' \mathbf{x}_{1,T_1^*+1} + \hat{\beta}_1' \mathbf{x}_{1,T_1^*} - \eta_1 - \phi_1 y_{1,T_1^*} \\
&\quad -\theta_1' \mathbf{x}_{1,T_1^*+1} - \beta_1' \mathbf{x}_{1,T_1^*} - \alpha_1 - \varepsilon_{1,T_1^*+1}) + \hat{\alpha}_{\mathrm{adj}}\big\}^2 \\
&= R_{T_1^*+1,1} + \mathrm{E}(\hat{\alpha}_{\mathrm{adj}}^2) - 2\,\mathrm{E}\Big\{(\hat{\eta}_1 + \hat{\phi}_1 y_{1,T_1^*} + \hat{\theta}_1' \mathbf{x}_{1,T_1^*+1} \\
&\quad +\hat{\beta}_1' \mathbf{x}_{1,T_1^*} - \eta_1 - \phi_1 y_{1,T_1^*} - \theta_1' \mathbf{x}_{1,T_1^*+1} - \beta_1' \mathbf{x}_{1,T_1^*} - \alpha_1 - \varepsilon_{1,T_1^*+1})\hat{\alpha}_{\mathrm{adj}}\Big\} \\
&= R_{T_1^*+1,1} + \mathrm{E}(\hat{\alpha}_{\mathrm{adj}}^2) - 2\,\mathrm{E}(\hat{\alpha}_{\mathrm{adj}})\,\mathrm{E}(\alpha_1) \\
&\quad - 2\,\mathrm{E}\Big\{(\hat{\eta}_1 + \hat{\phi}_1 y_{1,T_1^*} + \hat{\theta}_1' \mathbf{x}_{1,T_1^*+1} + \hat{\beta}_1' \mathbf{x}_{1,T_1^*} - \eta_1 - \phi_1 y_{1,T_1^*} - \theta_1' \mathbf{x}_{1,T_1^*+1} - \beta_1' \mathbf{x}_{1,T_1^*})\hat{\alpha}_{\mathrm{adj}}\Big\} \\
&= R_{T_1^*+1,1} + \mathrm{Var}(\hat{\alpha}_{\mathrm{adj}}) - \mu_\alpha^2 \\
&\quad - 2\,\mathrm{E}\Big(\hat{\eta}_1 + \hat{\phi}_1 y_{1,T_1^*} + \hat{\theta}_1' \mathbf{x}_{1,T_1^*+1} + \hat{\beta}_1' \mathbf{x}_{1,T_1^*} - \eta_1 - \phi_1 y_{1,T_1^*} - \theta_1' \mathbf{x}_{1,T_1^*+1} - \beta_1' \mathbf{x}_{1,T_1^*}\Big)\,\mu_\alpha,
\end{aligned}
$$

where

$$
\mathrm{E}(\hat{\alpha}_{\mathrm{adj}}) = \mathrm{E}\left\{\mathrm{E}(\hat{\alpha}_{\mathrm{adj}}|\mathcal{H})\right\} = \mathrm{E}\left(\frac{1}{n}\sum_{i=2}^{n+1}\alpha_i\right) = \mu_\alpha,
$$

and $\mathcal{H}$ defined in (3). Observe that

$$
\begin{aligned}
&\mathrm{E}\Big(\hat{\eta}_1 + \hat{\phi}_1 y_{1,T_1^*} + \hat{\theta}_1' \mathbf{x}_{1,T_1^*+1} + \hat{\beta}_1' \mathbf{x}_{1,T_1^*} - \eta_1 - \phi_1 y_{1,T_1^*} - \theta_1' \mathbf{x}_{1,T_1^*+1} - \beta_1' \mathbf{x}_{1,T_1^*}\Big) \\
&= \mathrm{E}\Big\{\mathrm{E}\Big(\hat{\eta}_1 + \hat{\phi}_1 y_{1,T_1^*} + \hat{\theta}_1' \mathbf{x}_{1,T_1^*+1} + \hat{\beta}_1' \mathbf{x}_{1,T_1^*} - \eta_1 - \phi_1 y_{1,T_1^*} - \theta_1' \mathbf{x}_{1,T_1^*+1} - \beta_1' \mathbf{x}_{1,T_1^*}\Big)\,|\mathcal{H}_1\Big\} \\
&= 0.
\end{aligned}
$$

Therefore,

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

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

When we estimate all of the parameters via OLS we can write $\mathrm{Var}(\hat{\alpha}_{\mathrm{adj}})$ as

$$
\begin{aligned}
\mathrm{Var}(\hat{\alpha}_{\mathrm{adj}}) &= \mathrm{E}\left\{\mathrm{Var}(\hat{\alpha}_{\mathrm{adj}}|\mathcal{H})\right\} + \mathrm{Var}\left\{\mathrm{E}(\hat{\alpha}_{\mathrm{adj}}|\mathcal{H})\right\} \\
&= \frac{1}{n^2}\sum_{i=2}^{n+1}\mathrm{E}\left\{\mathrm{Var}(\hat{\alpha}_i|\mathcal{H})\right\} + \frac{\sigma_\alpha^2}{n}.
\end{aligned} \tag{4}
$$

Propostion 1 states that Forecast 2 has lower forecast risk than Forecast 1 when $\mathrm{Var}(\hat{\alpha}_{\mathrm{adj}}) < \mu_\alpha^2$. Combining this result with (4), we see that Forecast 2 has lower forecast risk than Forecast 1 when

$$
\frac{1}{n^2}\sum_{i=2}^{n+1}\mathrm{E}\left\{\mathrm{Var}(\hat{\alpha}_i|\mathcal{H})\right\} + \frac{\sigma_\alpha^2}{n} < \mu_\alpha^2.
$$

Forecast 2 is preferable to Forecast 1 asymptotically in both T and n whenever $\mu_\alpha \neq 0$. In finite samples, Forecast 2 is preferable to Forecast 1 when the μ_α is large relative to its variability and overall regression variability.

References

John M Bates and Clive WJ Granger. The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969.

Richard Blundell and Stephen Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143, 1998.

Clive William John Granger and Paul Newbold. *Forecasting economic time series*. Academic Press, 2014.

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

Allan Timmermann. Forecast combinations. *Handbook of economic forecasting*, 1:135–196, 2006.