Prospective testing for the prevalence or transience of a shock effect before it occurs

Abstract

We develop a hypothesis testing procedure to prospectively test whether an anticipated shock is likely to be transient or permanent over a time horizon. We achieve this by borrowing knowledge from other time series that have undergone similar shocks for which post-shock outcomes are observed. These additional time series form a donor pool. For each of the time series in the donor pool we calculate a p-value corresponding to a hypothesis test on the relevance of the inclusion of shock-effect information in predicting the response over the time horizon. These p-values are then combined to form an aggregated p-value which guides one decision in determining whether the shock effect for the time series under study is expected to be prevalent or transient. This p-value can be computed before the shock-effect is observed in the time series under study provided one can form a suitable donor pool. Several simulated data examples, and two real data examples of forecasting Conoco Phillips stock price and .... are provided for verification and illustration.

1 Introduction

We provide forecasting methodology for assessing the lingering effect of an anticipated structural shock to a time series under study. We focus on the setting in which a structural shock has occurred and one desires a prediction for the post-shock response over a set time horizon H. Specific interest is in determining whether the shock is expected to be permanent or transient over H. Standard forecasting methods may not yield any guidance on the post-shock trajectories [**?**]. This is a general problem that has many real life applications. For example, one may acquire terrible or great news about a company and desire to determine whether that news is bound to impact the stock price of that company over a relevant time period. Companies may be interested in forecasting the demand of their products after they were involved in a brand crisis, but they only have recent sales data from pre-crises times. 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 post-shock trajectories arising from materially 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 testing method embraces ideas from forecast aggregation in the post-shock setting [**?**], forecast comparison [**??**], p-value combination, conditional forecasting [**??**], time series pooling using cross-sectional panel data [**??????**], forecasting with judgement and models [**??**], synthetic control methodology [**??**], expectation shocks [**???**].

2 Setting

We will suppose that a researcher has multivariate time series data $\mathbf{y}_{i,t}$, $t = 1, \ldots, T_i$ and $i = 1, \ldots, n+1$. We let $\mathbf{y}_{i,t} = (y_{i,t},\, \mathbf{x}_{i,t})$ where $y_{i,t}$ is a scalar response and $\mathbf{x}_{i,t}$ is a vector of covariates that are revealed to the analyst prior to the observation of $y_{1,t}$. Suppose that the analyst is interested in forecasting $y_{1,t}$, the first time series in the collection. We will suppose that each time series $\mathbf{y}_{i,t}$ undergoes a shock at time $T_i^* \leq T_i + 1$. To define an interesting setting, we will suppose that $T_1^* = T_1 + 1$, and $1 < T_i^* < T_i + 1$ for $i \geq 2$. We will suppose that $\mathbf{x}_{i,t=T_i^*}$ is observed before the shock takes effect on $y_{i,t=T_i^*}$.

We are interested in point forecasts $y_{i,t}^h$ at multiple horizons, $h = 1, \ldots, H$ with the aim of determining whether the shock has an effect on $y_{i,t}^h$. **?** provided a methodology for comparing forecasts jointly across all horizons of a forecast path, $h = 1, \ldots, H$. In our post-shock setting, we want to compare the forecasts

$$\hat{y}_{1,t}^{i,h} \text{ and } \hat{y}_{i,t}^{2,h}$$

where $y_{i,t}^{1,h}$ is the forecast for $y_{i,t}$ that accounts for the yet-to-be observed structural shock and is based on the information set $\mathcal{F}_{t-h}$, and $\hat{y}_{i,t}^{2,h}$ is defined similarly for the forecast that does not include any shock effect information. We will compare these forecasts in terms of their loss differential

$$\mathbf{d}_{i,t} = \text{Ł}_{i,t,1} - \text{Ł}_{i,t,2},$$

where $L_{i,t,j} \in \mathbb{R}^H$ has elements $L^h(y_{i,t},\, \hat{y}_{i,t}^{h,j})$, $j = 1, 2$, and L is a loss function. Hypothesis tests in **?** are with respect to $E(\mathbf{d}_{i,t}) = \mu_{i,t}$. Conditions for these tests require conditions of **?**.

We will be interested in $\mu_i = \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} \mu_{i,t}$.

Note: We need more formality for constructing $\hat{y}_{1,t}^{1,h}$. We could use the forecasts in **?** and then consider h-ahead methods after adjusting for the shock. Or we could consider aggregation approaches which average all post-shock responses of the series in the donor pool.

We will consider the average superior predictive ability (aSPA) to assess whether or not a shock is permanent or transitory. The aSPA investigates forecast comparisons based on their weighted average loss difference

$$\mu^{(AVG)} = \mathbf{w}^T \mu = \sum_{h=1}^{H} w_h \mu^h$$

with weights $\mathbf{w}$ that sum to one. Note that aSPA requires the user to take a stand on the relative importance of under-performance at one horizon against out-performance at another, and note that it is likely that $\mu^h > 0$ for h closer to 1 since the user expects that a structural shock will occur and the structural shock is taken into account by forecast 1.

2.1 Model setup

Note: We need to update the modeling setup. We should present a general modeling class which includes both the decay model and the permanent shift model.

Possible modeling approach

We now describe the assumed autoregressive models with random effects for which post-shock aggregated estimators are provided. The model $\mathcal{M}$ is defined as

$$\mathcal{M}\colon \begin{array}{l} y_{i,t} = \eta_i + \sum_{j=1}^{q_1} \phi_{i,j} y_{i,t-j} + \sum_{j=0}^{q_2-1} \theta_{i,j+1}' \mathbf{x}_{i,t-j} + \alpha_i D_{i,t} f(t) + \varepsilon_{i,t}, \\ \alpha_i = \mu_\alpha + \delta_i' \mathbf{x}_{i,T_i^*+1} + \tilde{\varepsilon}_i, \end{array} \tag{1}$$

where $D_{i,t} = I(t \geq T_i^* + 1)$, $I(\cdot)$ is the indicator function, $\mathbf{x}_{i,t} \in \mathbb{R}^p$ are fixed with $p \geq 1$, $f(t)$ is a bounded continuous function which does not cross zero and $\lim_{t\to\infty} g(t) = a \in \mathbb{R}$. Let $\phi_i = (\phi_{i,1}, \ldots, \phi_{i,q_1})'$,

$\theta_i = (\theta_{i,1}, \ldots, \theta_{i,q_2})'$, $\delta_i = (\delta_{i,1}, \ldots, \delta_{i,q_3})'$, and suppose that the regression coefficients in (1) have the following random effects structure:

$$\begin{aligned}
\eta_i &\overset{iid}{\sim} \mathcal{F}_\eta \text{ with } \mathrm{E}_{\mathcal{F}_\eta}(\eta_i) = 0, \mathrm{Var}_{\mathcal{F}_\eta}(\eta_i) = \sigma_\eta^2, \\
\phi_i &\overset{iid}{\sim} \mathcal{F}_\phi \text{ where } |\phi_{i,j}| < 1, \\
\theta_i &\overset{iid}{\sim} \mathcal{F}_\theta \text{ with } \mathrm{E}_{\mathcal{F}_\theta}(\theta_i) = \mu_\theta, \mathrm{Var}_{\mathcal{F}_\theta}(\theta_i) = \Sigma_\theta^2, \\
\delta_i &\overset{iid}{\sim} \mathcal{F}_\delta \text{ with } \mathrm{E}_{\mathcal{F}_\delta}(\delta_i) = \mu_\delta, \mathrm{Var}_{\mathcal{F}_\delta}(\delta_i) = \Sigma_\delta, \\
\varepsilon_{i,t} &\overset{iid}{\sim} \mathcal{F}_{\varepsilon_i} \text{ with } \mathrm{E}_{\mathcal{F}_{\varepsilon_i}}(\varepsilon_{i,t}) = 0, \mathrm{Var}_{\mathcal{F}_{\varepsilon_i}}(\varepsilon_{i,t}) = \sigma_i^2, \\
\tilde{\varepsilon}_i &\overset{iid}{\sim} \mathcal{F}_{\tilde{\varepsilon}} \text{ with } \mathrm{E}_{\mathcal{F}_{\tilde{\varepsilon}}}(\tilde{\varepsilon}_i) = 0, \mathrm{Var}_{\mathcal{F}_{\tilde{\varepsilon}}}(\tilde{\varepsilon}_i) = \sigma_\alpha^2, \\
\eta_i &\perp\!\!\!\perp \phi_i \perp\!\!\!\perp \theta_i \perp\!\!\!\perp \delta_i \perp\!\!\!\perp \varepsilon_{i,t} \perp\!\!\!\perp \tilde{\varepsilon}_{i,t}.
\end{aligned}$$

The model (1) with the above random effects structure is a generalization of both model formulations in **?**. **Need to carefully show**.

We further define the parameter sets

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

Alternative model formulation

We now describe the assumed autoregressive models with random effects for which post-shock aggregated estimators are provided. The model $\mathcal{M}$ is defined as

$$\mathcal{M}\colon \quad \begin{aligned}
& y_{i,t} = \eta_i + \textstyle\sum_{j=1}^{q_1} \phi_{i,j} y_{i,t-j} + \sum_{j=0}^{q_2-1} \theta_{i,j+1}' \mathbf{x}_{i,t-j} + f(\mathcal{F}_{i,t}, \alpha_i) D_{i,t} + \varepsilon_{i,t}, \\
& f(\mathcal{F}_{i,t}, \alpha_i) = \alpha_i g(t) + \textstyle\sum_{j=1}^{\tilde{q}_1} \tilde{\phi}_{i,j} y_{i,t-j} + \sum_{j=0}^{\tilde{q}_2-1} \tilde{\theta}_{i,j+1}' \mathbf{x}_{i,t-j}, \\
& \alpha_i = \mu_\alpha + \delta_i' \mathbf{x}_{i,T_i^*+1} + \tilde{\varepsilon}_i,
\end{aligned} \tag{3}$$

where $g(t)$ is a known or estimable bounded continuous function which does not cross zero and $\lim_{t\to\infty} g(t) = a \in \mathbb{R}$. Let $\phi_i = (\phi_{i,1}, \ldots, \phi_{i,q_1})'$, $\theta_i = (\theta_{i,1}, \ldots, \theta_{i,q_2})'$, $\tilde{\phi}_i = (\tilde{\phi}_{i,1}, \ldots, \tilde{\phi}_{i,\tilde{q}_{i,1}})'$, $\tilde{\theta}_i = (\tilde{\theta}_{i,1}, \ldots, \tilde{\theta}_{i,\tilde{q}_{i,2}})'$, $\delta_i = (\delta_{i,1}, \ldots, \delta_{i,q_3})'$, and suppose that the regression coefficients in (3) have the following hierarchical random effects structure:

$$\begin{aligned}
\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 } |\phi_{i,j}| < 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, \\
\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, \\
\varepsilon_{i,t} &\overset{iid}{\sim} \mathcal{F}_{\varepsilon_i} \text{ with } \mathrm{E}_{\mathcal{F}_{\varepsilon_i}}(\varepsilon_{i,t}) = 0, \mathrm{Var}_{\mathcal{F}_{\varepsilon_i}}(\varepsilon_{i,t}) = \sigma_i^2, \\
\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, \\
\tilde{\phi}_{i,j} &= \mu_{\tilde{\phi}_j} + \lambda_{i,j}, j = 1, \ldots, \tilde{q}_{i,1}, \text{ where } \lambda_{i,j} \overset{ind}{\sim} \mathcal{F}_\lambda(\mathbf{x}_{i,T_i^*}) \text{ with } |\mu_{\tilde{\phi}_j}| < 1 \text{ and } |\mu_{\tilde{\phi}_j} + \lambda_{i,j}| < 1, \\
\tilde{\theta}_{i,j} &= \theta_{i,j} + \gamma_{i,j}, j = 1, \ldots, \tilde{q}_{i,2}, \text{ where } \gamma_{i,j} \overset{iid}{\sim} \mathcal{F}_\gamma \text{ with } \mathrm{E}_{\mathcal{F}_\gamma}(\gamma_i) = \beta_o + \beta \mathbf{x}_{i,T_i^*+1}, \mathrm{Var}_{\mathcal{F}_\gamma}(\gamma_i) = \Sigma_\gamma^2, \\
\tilde{q}_{i,k} &\overset{ind}{\sim} \mathcal{F}_k(\mathbf{x}_{i,T_i^*}), \ k = 1, 2, \\
\eta_i &\perp\!\!\!\perp \phi_i \perp\!\!\!\perp \theta_i \perp\!\!\!\perp \tilde{\phi}_i \perp\!\!\!\perp \lambda_{i,j} \perp\!\!\!\perp \gamma_i \perp\!\!\!\perp \delta_i \perp\!\!\!\perp \tilde{q}_{i,k} \perp\!\!\!\perp \varepsilon_{i,t} \perp\!\!\!\perp \tilde{\varepsilon}_{i,t},
\end{aligned}$$

where the distribution $\mathcal{F}_k(\mathbf{x}_{i,T_i^*})$ is a discrete distribution conditional on $\mathbf{x}_{i,T_i^*}$, $\theta_{i,j} = 0$ for any $q_1 < j \leq \tilde{q}_{i,1}$, and $\phi_{i,j} = 0$ for any $q_2 < j \leq \tilde{q}_{i,2}$. The model (3) with the above random effects structure is a generalization of both model formulations in **?**. **Need to carefully show**. Note that for model (3) to be of use for post-shock forecasting, the variation in $\mathcal{F}_k(\mathbf{x}_{i,T_i^*})$ and $\mathcal{F}_\lambda(\mathbf{x}_{i,T_i^*})$ needs to be small relative to the signal captured in $\mathbf{x}_{i,T_i^*}$, and σ_α^2 and Σ_γ^2 needs to be small relative to $\mathbf{x}_{i,T_i^*}$.

2.2 Forecasting and testing for shock persistence

Note: Our forecast needs to be written with respect to our general model. Specifics can be given when we conduct our numerical examples.

In our post-shock setting we consider the following candidate forecasts:

$$\text{Forecast 1}:$$
$$\text{Forecast 2}:$$

We want to determine which forecast is appropriate over a horizon while the methods in **?** were only appropriate in the nowcasting setting in which prediction was only focused on the response immediately following the shock.

Proposition 1. *Let $p_i \sim \mathcal{D}$ be independent sequence of p-values for $i = 1, \ldots, n+1$ and some distribution $\mathcal{D}$, where p_1 is the p-value of the time series of interest. Let α denote the significance level. If $\mathbb{P}(p_1 \leq \alpha) \neq 0.5$, with probability one, the expected misclassification rate for voting in prevalence testing is*

$$\begin{cases} 1 - \mathbb{P}(p_1 \leq \alpha) & \textit{if } \mathbb{P}(p_1 \leq \alpha) > 0.5 \\ \mathbb{P}(p_1 \leq \alpha) & \textit{if } \mathbb{P}(p_1 \leq \alpha) < 0.5 \end{cases}.$$

Proof. Since $\mathbb{P}(p_1 \leq \alpha) \in \mathbb{R}$, by Strong Law of Large Numbers,

$$\frac{1}{n}\sum_{i=2}^{n+1} I(p_i \leq \alpha) \overset{a.s.}{\to} \mathbb{P}(p_i \leq \alpha) = \mathbb{P}(p_1 \leq \alpha),$$

which follows from the fact that p_i are i.i.d. Define

$$f\colon [0,1] \mapsto \{0,1\} \text{ with } f(x) = I(x \geq 0.5).$$

Let $C(f)$ denote the continuity set of f. Suppose that $\mathbb{P}(p_1 \leq \alpha) \neq 0.5$. In this case, notice that

$$\mathbb{P}(\mathbb{P}(p_1 \leq \alpha) \in C(f)) = 1.$$

By Slutsky's Theorem, we have

$$I\left\{\frac{1}{n}\sum_{i=2}^{n+1} I(p_i \leq \alpha) \geq 0.5\right\} \overset{a.s.}{\to} I\{\mathbb{P}(p_1 \leq \alpha) \geq 0.5\}.$$

Moreover, note that

$$\begin{aligned} \mathbb{E}\left\{|I\{\mathbb{P}(p_1 \leq \alpha) \geq 0.5\} - I(p_1 \leq \alpha)|\right\} &= \begin{cases} 1 - \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) > 0.5 \\ \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) < 0.5 \end{cases} \\ &\leq 0.5. \end{aligned}$$

That implies that with probability one,

$$\mathbb{E}\left\{\left|I\left\{\frac{1}{n}\sum_{i=2}^{n+1} I(p_i \leq \alpha) \geq 0.5\right\} - I(p_1 \leq \alpha)\right|\right\} = \begin{cases} 1 - \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) > 0.5 \\ \mathbb{P}(p_1 \leq \alpha) & \text{if } \mathbb{P}(p_1 \leq \alpha) < 0.5 \end{cases}.$$

That is, with probability one, the expected misclassification rate is as above. $\square$

Remark 1. Since indicator functions are bounded in $\mathcal{L}^2$, by Chebyshev-Rachman Strong Law of Large Numbers, the independence assumption can be relaxed to be that those p-values are uncorrelated.

3 Simulation Setup

Let n denote the donor pool size, p denote the number of covariates used, H denote the number of horizon used, T_i denote the length of time series to be evaluated for time series i, K_i denote the training sample size used for each forecasting time series i, T_i^* denote the time point just before the realization of the shock for time series i for $i = 1, \ldots, n+1$.

In this setting n, p, and H are pre-determined. $T_i, K_i \sim \text{Gamma}(15, 10)$. The total sample size for ith time series is $T_i + K_i + H$. T_i^* is randomly sampled from $\lceil \frac{1}{4} T_i \rceil + 1$ to $\lceil \frac{3}{4} T_i \rceil + K_i + H$. If $T_i, K_i < 90$, we force them to be 90. The adopted model for the data is as below:

$$
\begin{aligned}
y_{i,t} &= \eta_i + \phi_i y_{i,t-1} + \mathbf{x}_{i,t}\boldsymbol{\beta}_i + \alpha_i I(t > T_i^*) + \varepsilon_{i,t},\\
\alpha_i &= \mu_\alpha + \mathbf{x}_{i,T_i^*+1}\boldsymbol{\gamma}_i + \tilde{\varepsilon}_i,
\end{aligned}
$$

where

$$
\begin{aligned}
\phi_i &\sim \text{ indep. } U(0,1)\\
\eta_i &\sim \text{ indep. } \mathcal{N}(0,1)\\
\varepsilon_{i,t} &\sim \text{ indep. } \mathcal{N}(0,\sigma^2)\\
\tilde{\varepsilon}_i &\sim \text{ indep. } \mathcal{N}(0,\sigma_\alpha^2)\\
\boldsymbol{\gamma}_i &\sim \text{ indep. } \mathcal{N}(\mu_\gamma \mathbf{1}_p, \sigma_\gamma^2 \mathbf{I}_p)\\
\boldsymbol{\beta}_i &\sim \text{ indep. } \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p).
\end{aligned}
$$

Moreover, the elements of $\mathbf{x}_{i,t}$ are independently distributed as $\text{Gamma}(1,\delta)$.

Note that K_i is training sample size for time series i. Consider

$$
\begin{aligned}
K_i &\sim \lceil \text{Gamma}(a_K, b_K) \rceil\\
T_i &\sim \lceil \text{Gamma}(a_T, b_T) \rceil\\
T_i^* &\equiv \max\{T_i + 1, \lceil 0.5 \cdot (T_i + K_i + H) \rceil\},
\end{aligned}
$$

$K_i + H + T_i^* > T_i + K_i + H$

Then, we consider the following simulation setup

```
ns <- c(5, 10, 20, 40)
Tscale <- Kscale <- 1 / 2 # b_T, b_K
K.T.shape <- c(200, 400, 800, 1600) # for K_i and T_i
mu.gamma.delta <- 2 # mean for parameter vector of shock
sigma.delta.gamma <- 0.1 # sd for parameter vector of shock
sigma.alpha <- 0.05 # sd for shock noise
sigma <- 0.1 # sd for response noise
mu.alpha <- 50 # intercept for shock (relatively large)
H <- 8
ell <- 4
scale <- 2 # scale for covariates that follow Gamma distribution
```

$$
\begin{aligned}
y_{i,t} &= \eta_i + \phi_i y_{i,t-1} + \mathbf{x}_{i,t}\boldsymbol{\beta}_i + \xi_i \cdot I(t > T_i^*) + \varepsilon_{i,t},\\
\xi_i &= \alpha_i \cdot e^{-(t-T_i^*-1)}\\
\alpha_i &= \mu_\alpha + \mathbf{x}_{i,T_i^*+1}\boldsymbol{\gamma}_i + \tilde{\varepsilon}_i,
\end{aligned}
$$

References