Probability Paths and the Structure of Predictions over Time: Supplementary Information

1 Proofs

1.1 Proof of Theorem 1

Proof. Initially, by definition of Y_0, the observed y_0 can be expressed as

$$y_0 = \Phi\left(\frac{\gamma + \mathbb{E}[\sum_{i=1}^T Z_i]}{\sqrt{\mathrm{Var}(\sum_{i=1}^T Z_i)}}\right) = \Phi\left(\frac{\gamma}{\sqrt{\sum_{i,j} \mathbf{\Sigma}_{(i,j)}}}\right)$$

and γ can be uniquely identified as

$$\gamma = \Phi^{-1}(y_0)\sqrt{\sum_{i,j} \mathbf{\Sigma}_{(i,j)}}.$$

Similarly, definition of Y_t gives

$$Y_t = 1 - \Phi\left(\frac{-\gamma - \sum_{i=1}^{t-1} Z_i - Z_t - \bar{\mu}_t}{\bar{\sigma}_t}\right).$$

By symmetry, we have

$$Y_t = \Phi\left(\frac{\gamma + \sum_{i=1}^{t-1} Z_i + Z_t + \bar{\mu}_t}{\bar{\sigma}_t}\right)$$

$$\Phi^{-1}(Y_t) = \frac{\gamma + \sum_{i=1}^{t-1} Z_i + Z_t + \bar{\mu}_t}{\bar{\sigma}_t}, \tag{1}$$

where $\Phi(\cdot)$ is the standard normal cumulative density function (CDF); $\bar{\mu}_t$ and $\bar{\sigma}_t$ are respectively the mean and standard deviation of the Gaussian variable $\sum_{i=t+1}^T Z_i$ conditioned on previous latent information variable values.

Now we show how to obtain the defining parameters $\bar{\mu}_t$ and $\bar{\sigma}_t$ for the conditional distribution of $(\sum_{i=t+1}^T Z_i \mid Z_1 = z_1, \ldots, Z_t = z_t)$.

At time t, given the realization of the latent information variables $z_1, \ldots, z_t$, the remaining ones will follow a multivariate Gaussian distribution:

$$(Z_{t+1}, \ldots, Z_T \mid Z_1 = z_1, \ldots, Z_t = z_t) \sim \mathcal{N}(\mu^t, \mathbf{\Sigma}^t) \tag{2}$$

with

$$\mu^t = \mathbf{\Sigma}_{21}^t \left(\mathbf{\Sigma}_{11}^t\right)^{-1} [z_1, \ldots, z_t]^T$$
$$\mathbf{\Sigma}^t = \mathbf{\Sigma}_{22}^t - \mathbf{\Sigma}_{21}^t \left(\mathbf{\Sigma}_{11}^t\right)^{-1} \mathbf{\Sigma}_{12}^t.$$

The terms μ^t and $\mathbf{\Sigma}^t$ are simply the conditional mean and variance of the multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ when conditioned on the first t latent information variables.

Then the conditional sum $(\sum_{i=t+1}^{T} Z_i \mid Z_1 = z_1, \ldots, Z_t = z_t)$ will follow a Gaussian distribution $\mathcal{N}(\bar{\mu}_t, \bar{\sigma}_t^2)$. Let $\mathbf{a}^t$ be $\mathbf{1}^T \mathbf{\Sigma}_{21}^t \left(\mathbf{\Sigma}_{11}^t\right)^{-1}$, we have mean $\bar{\mu}_t$ and variance $\bar{\sigma}_t^2$ being

$$\begin{aligned}\bar{\mu}_t &= \mathbf{1}^T \mu^t \\ &= \mathbf{1}^T \mathbf{\Sigma}_{21}^t \left(\mathbf{\Sigma}_{11}^t\right)^{-1} [z_1, \ldots, z_t]^T \\ &= \mathbf{a}^t [z_1, \ldots, z_t]^T = \sum_{i=1}^{t} \mathbf{a}_{(i)}^t z_i \\ \bar{\sigma}_t^2 &= \sum_{i,\, j} \mathbf{\Sigma}_{(i,j)}^t,\end{aligned}$$

where $\mathbf{\Sigma}_{(i,j)}^t$ is the element at the i-th row, j-th column in $\mathbf{\Sigma}^t$.

Once we have observed y_t and identified $z_1, \ldots, z_{t-1}$, by substituting and rearranging Eq. (1), we can uniquely identify Z_t:

$$Z_t = z_t = \frac{\bar{\sigma}_t \Phi^{-1}(y_t) - \sum_{i=1}^{t-1}(1 + \mathbf{a}_{(i)}^t) z_i - \gamma}{1 + \mathbf{a}_{(t)}^t}.$$

Therefore, given $\mathbf{\Sigma}$ and γ, once we have observed $y_1, \ldots, y_t$, we can uniquely identify $z_1, \ldots, z_t$. As a result, $(\Phi^{-1}(Y_t) \mid Y_{t-1} = y_{t-1}, \ldots, Y_1 = y_1)$ is equivalent of $(\Phi^{-1}(Y_t) \mid Z_{t-1} = z_{t-1}, \ldots, Z_1 = z_1)$.

Now we are going to find the conditional distribution for $(\Phi^{-1}(Y_t) \mid Z_{t-1} = z_{t-1}, \ldots, Z_1 = z_1)$.

As shown in Eq. (2), when conditioning on the first t latent information variables Zs, the remaining ones follow distribution $\mathcal{N}(\mu^t, \mathbf{\Sigma}^t)$. Then when conditioned on $(Z_{t-1} = z_{t-1}, \ldots, Z_1 = z_1)$, the mean and variance of the conditional marginal distribution of Z_t are simply the first elements in μ^{t-1} and $\mathbf{\Sigma}^{t-1}$, namely:

$$(Z_t \mid Z_{t-1} = z_{t-1}, \ldots, Z_1 = z_1) \sim \mathcal{N}(\mu_{(1)}^{t-1}, \mathbf{\Sigma}_{(1,1)}^{t-1}).$$

Substituting $\bar{\mu}_t$, $\bar{\sigma}_t^2$, and replacing the conditioned $Z_{t-1}, \ldots, Z_1$ with $z_{t-1}, \ldots, z_1$ in Eq. (1), we can obtain the conditional distribution of $\Phi^{-1}(Y_t)$, which is a linear transformation of $(Z_t \mid Z_{t-1} = z_{t-1}, \ldots, Z_1 = z_1)$:

$$(\Phi^{-1}(Y_t) \mid Z_{t-1} = z_{t-1}, \ldots, Z_1 = z_1) \sim \mathcal{N}(\tilde{\mu}_t, \tilde{\sigma}_t^2),$$

with

$$\begin{aligned}\tilde{\mu}_t &= \frac{\gamma + \sum_{i=1}^{t-1}(1 + \mathbf{a}_{(i)}^t) z_i + (1 + \mathbf{a}_{(t)}^t)\mu_{(1)}^{t-1}}{\bar{\sigma}_t} \\ \tilde{\sigma}_t &= \frac{\sqrt{\mathbf{\Sigma}_{(1,1)}^{t-1}}(1 + \mathbf{a}_{(t)}^t)}{\bar{\sigma}_t}.\end{aligned}$$

Finally, by applying change-of-variable trick, we are able to write out the conditional likelihood $P(Y_t = y_t \mid y_{t-1}, \ldots y_1; \mathbf{\Sigma}, \gamma)$ in terms of $\Phi^{-1}(y_t)$ for $0 < t < T$ as

$$\begin{aligned}&P(Y_t = y_t \mid y_{t-1}, \ldots y_1; \mathbf{\Sigma}, \gamma) \\ &\quad = P(Y_t = y_t \mid z_{t-1}, \ldots z_1; \mathbf{\Sigma}, \gamma) \\ &\quad = P(\Phi^{-1}(Y_t) = \Phi^{-1}(y_t) \mid z_{t-1}, \ldots z_1; \mathbf{\Sigma}, \gamma) \times \left| \frac{\partial \Phi^{-1}(y)}{\partial y}\bigg|_{y=y_t} \right| \\ &\quad = \frac{\varphi(\Phi^{-1}(y_t); \tilde{\mu}_t, \tilde{\sigma}_t^2)}{\varphi(\Phi^{-1}(y_t))}, \qquad (3)\end{aligned}$$

where $\varphi(\cdot; \tilde{\mu}_t, \tilde{\sigma}_t^2)$ is the PDF of a normal distribution with mean $\tilde{\mu}_t$ and variance $\tilde{\sigma}_t^2$ and $\varphi(\cdot)$ is the standard normal PDF. When $t = T$, we have $P(Y_T = y_T \mid y_{T-1}, \ldots y_1; \mathbf{\Sigma}, \gamma) = P(Y_T = y_T \mid y_{T-1}) = y_{T-1}^{y_T}(1 - y_{T-1})^{(1-y_T)}$ as it follows a Bernoulli distribution with $p = y_{T-1}$ by definition. Multiplying $P(Y_t = y_t \mid y_{t-1}, \ldots y_1; \mathbf{\Sigma}, \gamma)$ for all $0 < t \leq T$ gives the likelihood of the path $y_1, \ldots, y_T$. With expansion of normal PDFs, log transformation, and some rearrangement of constants, we have the expression for log PDF of $y_1, \ldots, y_T$ as presented in Theorem 1. □

1.2 Proof of Corollary 1

Proof. If $\mathbf{\Sigma}$ is diagonal, we have $\mathbf{\Sigma}_{21}^t = \mathbf{\Sigma}_{12}^t = \mathbf{0}$ for $t = 1, \ldots, T$, indicating $\mathbf{\Sigma}^t = \mathbf{\Sigma}_{22}^t$, $\mu_t = \mathbf{0}$, and $\mathbf{a}^t = \mathbf{0}$. We have,

$$
\begin{aligned}
z_t &= \frac{\bar{\sigma}_t \Phi^{-1}(y_t) - \sum_{i=1}^{t-1}(1 + \mathbf{a}_{(i)}^t) z_i - \gamma}{1 + \mathbf{a}_{(t)}^t} \\
&= \sqrt{\sum_{i=t+1}^{T} \sigma_i^2 \Phi^{-1}(y_t)} - \sum_{i=1}^{t-1} z_i - \gamma.
\end{aligned}
$$

Hence,

$$
\gamma + \sum_{i=1}^{t} z_i = \sqrt{\sum_{i=t+1}^{T} \sigma_i^2 \Phi^{-1}(y_t)}.
$$

Then

$$
\begin{aligned}
\tilde{\mu}_t &= \frac{\gamma + \sum_{i=1}^{t-1} z_i}{\sqrt{\sum_{i=t+1}^{T} \sigma_i^2}} = \Phi^{-1}(y_{t-1}) \sqrt{\frac{\sum_{i=t}^{T} \sigma_i^2}{\sum_{i=t+1}^{T} \sigma_i^2}} \\
\tilde{\sigma}_t &= \frac{\sqrt{\mathbf{\Sigma}_{(1,1)}^{t-1}}}{\sqrt{\sum_{i,j} \mathbf{\Sigma}_{(i,j)}^t}} = \frac{\sigma_t}{\sqrt{\sum_{i=t}^{T} \sigma_i^2}}.
\end{aligned}
$$

□

2 Metrics

2.1 Realized volatility MSE

The realized volatility $\text{MSE}_{t,w}^{\text{vol}}$ between time t and $t + w$ is calculated as

$$
\text{MSE}_{t,w}^{\text{vol}} = \frac{1}{N} \sum_{i=1}^{N} (\hat{V}_{i,t} - V_{i,t})^2, \tag{4}
$$

where

- N is the total number of probability paths we evaluate with;
- $r_{i,k}$ is the change in *observed* probability path between time $k-1$ and k for observation i;
- $\hat{r}_{i,j,k}$ is the change in *sample* probability path j between time $k-1$ and k for observation i;
- w is the size of the sliding window (i.e., 5 time steps in our case);
- m is the number of posterior paths we generate for each observation;
- $\hat{V}_{i,t} = \frac{1}{m} \sum_{j=1}^{m} \sum_{k=t+1}^{t+w} \hat{r}_{i,j,k}^2$ is the *estimated realized volatility* for probability path i between time t and $t + w$;
- $V_{i,t} = \sum_{k=t+1}^{t+w} r_{i,k}^2$ is the *ground truth realized volatility* for probability path i between time t and $t + w$.