IN-KOO CHO, KENNETH KASA: GRESHAM'S LAW OF MODEL AVERAGING

Discussion by Jaroslav Borovička and Anmol Bhandari

February 2016

- Abstract from persistent movements in fundamentals:

$$\begin{aligned} dx_t &= Ax_t dt + BdW_t \\ dy_t &= Dx_t dt + GdW_t \end{aligned}$$

- Abstract from persistent movements in fundamentals:

$$\begin{aligned} dx_t &= Ax_t dt + BdW_t \\ dy_t &= Dx_t dt + GdW_t \end{aligned}$$

- Two models for learning about β (here x_t):
    - **Model 0**: $A = B = 0$.
    - **Model 1**: $A = 0$, $B > 0$.

- Abstract from persistent movements in fundamentals:

$$
\begin{aligned}
dx_t &= Ax_t dt + BdW_t \\
dy_t &= Dx_t dt + GdW_t
\end{aligned}
$$

- Two models for learning about β (here x_t):
  - **Model 0**: $A = B = 0$.
  - **Model 1**: $A = 0$, $B > 0$.
- Standard solution:

$$
\begin{aligned}
d\bar{x}_t &= A\bar{x}_t dt + K_t \left(dy_t - D\bar{x}_t dt \right) \\
K_t &= \left[BG' + \boldsymbol{\Sigma}_t D' \right] \left(GG' \right)^{-1} \\
\frac{d\boldsymbol{\Sigma}_t}{dt} &= A\boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t A' + BB' - K_t GG' K_t'
\end{aligned}
$$

- Index (almost) everything by i

$$\begin{aligned} dx_t(i) &= B(i)\, dW_t(i) \\ dy_t &= D(i)\, x_t(i)\, dt + G dW_t(i) \end{aligned}$$

  - G has to stay the same for both models, otherwise immediately learnable

- Index (almost) everything by i

$$
\begin{aligned}
dx_t(i) &= B(i)\, dW_t(i) \\
dy_t &= D(i)\, x_t(i)\, dt + G dW_t(i)
\end{aligned}
$$

  - G has to stay the same for both models, otherwise immediately learnable
- **Model 0**: $A = B = 0$.
  - $\Sigma_t(0) \to 0$
- **Model 1**: $A = 0$, $B > 0$.
  - $\Sigma_t(1) \to \Sigma$.

- Index (almost) everything by i

$$\begin{aligned} dx_t(i) &= B(i)\, dW_t(i) \\ dy_t &= D(i)\, x_t(i)\, dt + G dW_t(i) \end{aligned}$$

  - G has to stay the same for both models, otherwise immediately learnable
- **Model 0**: $A = B = 0$.
  - $\boldsymbol{\Sigma}_t(0) \to 0$
- **Model 1**: $A = 0$, $B > 0$.
  - $\boldsymbol{\Sigma}_t(1) \to \boldsymbol{\Sigma}$.
- The Kalman-filter solution conditional on a model is

$$\begin{aligned} d\bar{x}_t(i) &= K_t(\boldsymbol{\Sigma}_t(i))\left[dy_t - D(i)\, \bar{x}_t(i)\, dt\right] \\ K_t(\boldsymbol{\Sigma}_t(i)) &= \left[B(i)\, G' + \boldsymbol{\Sigma}_t(i)\, D'(i)\right]\left[GG'\right]^{-1} \\ \frac{d\boldsymbol{\Sigma}_t(i)}{dt} &= \left(B(i) - K_t(\boldsymbol{\Sigma}_t(i))\, G\right)\left(B(i) - K_t(\boldsymbol{\Sigma}_t(i))\, G\right)' \end{aligned}$$

- Model indicator: $i \in \{0, 1\}$
- **Posterior probability** of $i = 1$ given $\mathcal{H}_t$

$$\bar{\imath}_t = E\left[i \mid \mathcal{H}_t\right]$$

- Model indicator: $i \in \{0, 1\}$
- **Posterior probability** of $i = 1$ given $\mathcal{H}_t$

$$\bar{\imath}_t = E\left[i \mid \mathcal{H}_t\right]$$

- By **Bayes rule** (prior $\bar{\imath}_0$, likelihood $l_t(i)$)

$$\bar{\imath}_t = \frac{\exp\left(l_t(1)\right)\bar{\imath}_0}{\exp\left(l_t(1)\right)\bar{\imath}_0 + \exp\left(l_t(0)\right)\left(1 - \bar{\imath}_0\right)}$$

- Model indicator: $i \in \{0,1\}$
- **Posterior probability** of $i = 1$ given $\mathcal{H}_t$

$$\bar{\imath}_t = E\left[i \mid \mathcal{H}_t\right]$$

- By **Bayes rule** (prior $\bar{\imath}_0$, likelihood $l_t\left(i\right)$)

$$\bar{\imath}_t = \frac{\exp\left(l_t\left(1\right)\right)\bar{\imath}_0}{\exp\left(l_t\left(1\right)\right)\bar{\imath}_0 + \exp\left(l_t\left(0\right)\right)\left(1-\bar{\imath}_0\right)}$$

- Derive the **law of motion** for $\bar{\imath}_t$ (innovations representation)

$$\begin{aligned} d\bar{\imath}_t &= \bar{\imath}_t\left(1-\bar{\imath}_t\right)\left[\left(D\left(1\right)\bar{x}_t\left(1\right) - D\left(0\right)\bar{x}_t\left(0\right)\right)'\left(GG'\right)^{-1}\right]\cdot \\ &\quad \cdot \underbrace{\left\{dy_t - \left[\bar{\imath}_t D\left(1\right)\bar{x}_t\left(1\right)dt + \left(1-\bar{\imath}_t\right)D\left(0\right)\bar{x}_t\left(0\right)dt\right]\right\}}_{\textit{innovation}} \end{aligned}$$

$$
\begin{aligned}
d\bar{\imath}_t &= \bar{\imath}_t \left(1-\bar{\imath}_t\right) \left[\left(D\left(1\right)\bar{x}_t\left(1\right)-D\left(0\right)\bar{x}_t\left(0\right)\right)' \left(GG'\right)^{-1}\right] \cdot \\
&\quad \cdot \left\{dy_t - \left[\bar{\imath}_t D\left(1\right)\bar{x}_t\left(1\right)dt + \left(1-\bar{\imath}_t\right)D\left(0\right)\bar{x}_t\left(0\right)dt\right]\right\}
\end{aligned}
$$

- $\bar{\imath}_t \in (0,1)$ with unattainable boundaries

$$\begin{aligned} d\bar{\imath}_t &= \bar{\imath}_t \left(1 - \bar{\imath}_t\right) \left[\left(D\left(1\right) \bar{x}_t\left(1\right) - D\left(0\right) \bar{x}_t\left(0\right)\right)' \left(GG'\right)^{-1} \right] \cdot \\ &\quad \cdot \left\{ dy_t - \left[\bar{\imath}_t D\left(1\right) \bar{x}_t\left(1\right) dt + \left(1 - \bar{\imath}_t\right) D\left(0\right) \bar{x}_t\left(0\right) dt\right] \right\} \end{aligned}$$

- $\bar{\imath}_t \in (0, 1)$ with unattainable boundaries
- Bounded martingale under agent's beliefs: **convergence**
  - Agent believes she will ultimately learn the probability of each of the models

$$
\begin{aligned}
d\bar{\imath}_t \quad = \quad & \bar{\imath}_t\left(1-\bar{\imath}_t\right)\left[\left(D\left(1\right)\bar{x}_t\left(1\right)-D\left(0\right)\bar{x}_t\left(0\right)\right)'\left(GG'\right)^{-1}\right]\cdot \\
& \cdot\left\{dy_t-\left[\bar{\imath}_t D\left(1\right)\bar{x}_t\left(1\right)dt+\left(1-\bar{\imath}_t\right)D\left(0\right)\bar{x}_t\left(0\right)dt\right]\right\}
\end{aligned}
$$

- $\bar{\imath}_t \in (0,1)$ with unattainable boundaries
- Bounded martingale under agent's beliefs: **convergence**
  - Agent believes she will ultimately learn the probability of each of the models
  - But what about the data-generating measure?

$$
\begin{aligned}
d\bar{\imath}_t &= \bar{\imath}_t \left(1 - \bar{\imath}_t\right) \left[\left(D\left(1\right) \bar{x}_t\left(1\right) - D\left(0\right) \bar{x}_t\left(0\right) \right)' \left(GG'\right)^{-1} \right] \cdot \\
&\quad \cdot \left\{ dy_t - \left[\bar{\imath}_t D\left(1\right) \bar{x}_t\left(1\right) dt + \left(1 - \bar{\imath}_t\right) D\left(0\right) \bar{x}_t\left(0\right) dt \right] \right\}
\end{aligned}
$$

- $\bar{\imath}_t \in (0, 1)$ with unattainable boundaries
- Bounded martingale under agent's beliefs: **convergence**
  - Agent believes she will ultimately learn the probability of each of the models
  - But what about the data-generating measure?
- 'Fast' dynamics in the center of $(0, 1)$, 'slow' close to the boundaries
  - The process spends most of the time close to the boundaries

- Under the data-generating measure, the model generates complicated dynamics.

- Under the data-generating measure, the model generates complicated dynamics.
- But we know how to analyze the dynamics c**onditional on each of the models.**

- Under the data-generating measure, the model generates complicated dynamics.
- But we know how to analyze the dynamics c**onditional on each of the models.**
- Close to the boundaries, this is enough.
  - Consider an approximation around $\bar{\imath}_t \approx 0$ (almost certainly believe in model 0)

$$d\bar{\imath}_t = \bar{\imath}_t \left[\left(D\left(1\right) \bar{x}_t\left(1\right) - D\left(0\right) \bar{x}_t\left(0\right) \right)' \left(GG' \right)^{-1} \right] \cdot \left\{ dy_t - D\left(0\right) \bar{x}_t\left(0\right) dt \right\}$$

- Under the data-generating measure, the model generates complicated dynamics.
- But we know how to analyze the dynamics c**onditional on each of the models.**
- Close to the boundaries, this is enough.
    - Consider an approximation around $\bar{\imath}_t \approx 0$ (almost certainly believe in model 0)

$$d\bar{\imath}_t = \bar{\imath}_t \left[\left(D(1)\,\bar{x}_t(1) - D(0)\,\bar{x}_t(0) \right)' \left(GG' \right)^{-1} \right] \cdot \{ dy_t - D(0)\,\bar{x}_t(0)\,dt \}$$

- How can we make this rigorous? Time scales

- $\bar{\imath}_t$, $\bar{x}_t(0)$, $\bar{x}_t(1)$ evolve at different '**speeds**'.
- $\bar{\imath}_t$ evolves more slowly than $\bar{x}_t(i)$

- $\bar{\imath}_t$, $\bar{x}_t(0)$, $\bar{x}_t(1)$ evolve at different '**speeds**'.
- $\bar{\imath}_t$ evolves more slowly than $\bar{x}_t(i)$
  - Consider the boundary behavior of $\log \bar{\imath}_t$ around $\bar{\imath}_t \approx 0$.

$$\begin{aligned} d \log \bar{\imath}_t \quad = \quad & -\frac{1}{2} \left| D(1)\,\bar{x}_t(1) - D(0)\,\bar{x}_t(0) \right|^2 dt + \\ & + \left[\left(D(1)\,\bar{x}_t(1) - D(0)\,\bar{x}_t(0) \right)' \left(GG' \right)^{-1} \right] \cdot \left\{ dy_t - D(0)\,\bar{x}_t(0)\,dt \right\} \end{aligned}$$

- $\bar{\imath}_t$, $\bar{x}_t(0)$, $\bar{x}_t(1)$ evolve at different '**speeds**'.
- $\bar{\imath}_t$ evolves more slowly than $\bar{x}_t(i)$
  - Consider the boundary behavior of $\log \bar{\imath}_t$ around $\bar{\imath}_t \approx 0$.

$$
\begin{aligned}
d \log \bar{\imath}_t \quad = \quad & -\frac{1}{2} \left| D(1)\,\bar{x}_t(1) - D(0)\,\bar{x}_t(0) \right|^2 dt + \\
& + \left[\left(D(1)\,\bar{x}_t(1) - D(0)\,\bar{x}_t(0) \right)' \left(GG' \right)^{-1} \right] \cdot \left\{ dy_t - D(0)\,\bar{x}_t(0)\,dt \right\}
\end{aligned}
$$

  - If the equilibrium is self-confirming, then $\bar{\imath}_t$ is attracted to zero.

- $\bar{\imath}_t$, $\bar{x}_t(0)$, $\bar{x}_t(1)$ evolve at different '**speeds**'.
- $\bar{\imath}_t$ evolves more slowly than $\bar{x}_t(i)$
  - Consider the boundary behavior of $\log \bar{\imath}_t$ around $\bar{\imath}_t \approx 0$.

$$\begin{aligned} d \log \bar{\imath}_t &= -\frac{1}{2} \left| D(1)\bar{x}_t(1) - D(0)\bar{x}_t(0) \right|^2 dt + \\ &+ \left[\left(D(1)\bar{x}_t(1) - D(0)\bar{x}_t(0) \right)' \left(GG' \right)^{-1} \right] \cdot \left\{ dy_t - D(0)\bar{x}_t(0)\, dt \right\} \end{aligned}$$

  - If the equilibrium is self-confirming, then $\bar{\imath}_t$ is attracted to zero.
  - Bounded drift and variance $\implies$ it can take arbitrarily long to recover from $\bar{\imath}_t \approx 0$.

- $\bar{\imath}_t$, $\bar{x}_t(0)$, $\bar{x}_t(1)$ evolve at different '**speeds**'.
- $\bar{\imath}_t$ evolves more slowly than $\bar{x}_t(i)$
  - Consider the boundary behavior of $\log \bar{\imath}_t$ around $\bar{\imath}_t \approx 0$.

$$\begin{aligned} d\log \bar{\imath}_t \quad = \quad & -\frac{1}{2}\left|D(1)\bar{x}_t(1) - D(0)\bar{x}_t(0)\right|^2 dt + \\ & + \left[\left(D(1)\bar{x}_t(1) - D(0)\bar{x}_t(0)\right)'\left(GG'\right)^{-1}\right] \cdot \left\{dy_t - D(0)\bar{x}_t(0)\,dt\right\} \end{aligned}$$

  - If the equilibrium is self-confirming, then $\bar{\imath}_t$ is attracted to zero.
  - Bounded drift and variance $\implies$ it can take arbitrarily long to recover from $\bar{\imath}_t \approx 0$.
  - $\bar{x}_t(i)$ can meanwhile trace out a sufficiently large set

The process spends more time close to $\bar{\imath}_t \approx 1$ than $\bar{\imath}_t \approx 0$.

The process spends more time close to $\bar{\imath}_t \approx 1$ than $\bar{\imath}_t \approx 0$.

Intuition: Imagine t large.

- Consider $\bar{\imath}_t \approx 0$. An unusal realization of dy_t contradicts model 0 since $\bar{x}_t(0)$ cannot move.

The process spends more time close to $\bar{\imath}_t \approx 1$ than $\bar{\imath}_t \approx 0$.

Intuition: Imagine t large.

- Consider $\bar{\imath}_t \approx 0$. An unusal realization of dy_t contradicts model 0 since $\bar{x}_t\,(0)$ cannot move.
  - Model 1 adjusts $\bar{x}_t\,(1)$ and now looks more likely $\implies$ $\bar{\imath}_t$ increases.

The process spends more time close to $\bar{\imath}_t \approx 1$ than $\bar{\imath}_t \approx 0$.

Intuition: Imagine t large.

- Consider $\bar{\imath}_t \approx 0$. An unusal realization of dy_t contradicts model 0 since $\bar{x}_t\,(0)$ cannot move.
  - Model 1 adjusts $\bar{x}_t\,(1)$ and now looks more likely $\implies$ $\bar{\imath}_t$ increases.
- Consider $\bar{\imath}_t \approx 1$. An unusal realization of dy_t can be rationalized by moving $\bar{x}_t\,(1)$
  - Self-confirmation in effect.

- In full information equilibrium, $\bar{\imath}_t = 0$ and $\bar{x}_t(0) = x$.

- In full information equilibrium, $\bar{\imath}_t = 0$ and $\bar{x}_t(0) = x$.
- But once we acknowledge incomplete information and the need for learning, the full information equilibrium is just an elusive dream.

WHY "GRESHAM'S" LAW?

- In full information equilibrium, $\bar{\imath}_t = 0$ and $\bar{x}_t(0) = x$.
- But once we acknowledge incomplete information and the need for learning, the full information equilibrium is just an elusive dream.
- Since the TVP model is self-confirming, there is nothing '**Gresham-like**' about it.

WHY "GRESHAM'S" LAW?

- In full information equilibrium, $\bar{\imath}_t = 0$ and $\bar{x}_t(0) = x$.
- But once we acknowledge incomplete information and the need for learning, the full information equilibrium is just an elusive dream.
- Since the TVP model is self-confirming, there is nothing '**Gresham-like**' about it.
- An omniscient entity may want the agents forget the TVP model but such entities are hard to come by.