Standard Signal Extraction

$(y, x)$ are scalars

$y_t = x_t + \varepsilon_t$

$x_t = \phi x_{t-1} + e_t$

$x$: signal

$y$: measurement

$\varepsilon$: measurement error
An example with $\sigma_e = 1$, $\sigma_e = 2$, $\phi = 0.9$ and $T = 200$ follows.
EX1.1: \( y(t) \) (Grey) and \( x(t) \) (Red)
EX1.2: $y(t)$ (Black), $x(t)$ (Red), $x(t/t)$ (Green)

(could also compute $x_{t|T}$)

387

Fig. 1. GDP and unemployment data. GDP_E and GDP_I are in growth rates and U_t is in changes. All are measured in annualized percent.
\[
\begin{bmatrix}
GDP_{Et} \\
GDP_{It}
\end{bmatrix} = 
\begin{bmatrix}
1 \\
1
\end{bmatrix} GDP_t + 
\begin{bmatrix}
\varepsilon_{Et} \\
\varepsilon_{It}
\end{bmatrix}
\]

\[GDP_t = \alpha + \rho GDP_{t-1} + \varepsilon_{Gt}\]

\[
\text{var} \begin{bmatrix}
\varepsilon_g \\
\varepsilon_E \\
\varepsilon_I
\end{bmatrix} = \Sigma = 
\begin{bmatrix}
\sigma_{GG} & 0 & 0 \\
0 & \sigma_{EE} & \sigma_{EI} \\
0 & \sigma_{EI} & \sigma_{II}
\end{bmatrix}
\]
Results:

For the 2-equation model with $\Sigma$ block-diagonal, we have

$$GDP_t = 3.06 (1 - 0.62) + 0.62 GDP_{t-1} + \epsilon_t, \quad (12)$$

$$\Sigma = \begin{bmatrix}
5.17 & 0 & 0 \\
0 & 3.86 & 1.43 \\
0 & 1.43 & 2.70
\end{bmatrix}.$$

$$\begin{bmatrix}
[2.77, 3.34] \\
[4.39, 5.95] \\
[3.34, 4.48] \\
[96, 1.95] \\
[0.96, 1.95] \\
[2.25, 3.22]
\end{bmatrix} \quad (13)$$
Fig. 3. GDP sample paths, 1960Q1–2011Q4. In each panel we show the sample path of \( GDP_M \) (light color) together with posterior interquartile range with shading and we show one of the competitor series (dark color). For \( GDP_M \) we use our benchmark estimate from the 2-equation model with \( \zeta = 0.80 \).
Figure 4: GDP Sample Paths, 2007Q1-2009Q4

Notes: In each panel we show the sample path of GDP\(_M\) in red together with a light-red posterior interquartile range, and we show one of the competitor series in black. For GDP\(_M\) we use our benchmark estimate from the 2-equation model with \(\beta = 0.80\).

First consider Figure 5. Across measurement-error models M, GDP\(_M\) is robustly more serially correlated than both GDP\(_E\) and GDP\(_I\), and also has small innovation variance. Hence most of our models achieve closely-matching unconditional variances, but they are composed of very different underlying (\(\alpha, \beta\)) values from those corresponding to GDP\(_E\).

GDP\(_M\) has smaller shock volatility, but much more shock persistence – roughly double that of GDP\(_E\) (\(\alpha\) of roughly 0.60 for GDP\(_M\) vs. 0.30 for GDP\(_E\)).

Now consider Table 1. The various GDP\(_M\) series are all less volatile than each of GDP\(_E\), GDP\(_I\) and GDP\(_F\), and also more skewed left. Most noticeably, the GDP\(_M\) series are much more strongly serially correlated than the GDP\(_E\), GDP\(_I\) and GDP\(_F\) series, and with smaller innovation variances. This translates into much higher predictive \(R^2\)'s for GDP\(_M\). Indeed GDP\(_M\) is twice as predictable as GDP\(_I\) or GDP\(_F\), which itself is predictably as GDP\(_E\).
Example: Missing Data

Scalar $y_t$ is generated by the ARMA(1,1) model

$$y_t = \phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$$

Observations available for $t = 1:45, 50:90, 100:130, 140:170: 180:200$

How can the likelihood be formed? How can the missing values be estimated?
State-space form of model:

\[
\begin{bmatrix}
\xi_t \\
y_t \\
e_t
\end{bmatrix},
F = \begin{bmatrix}
\phi & -\theta \\
0 & 0
\end{bmatrix},
\nu_t = \begin{bmatrix}
e_t \\
e_t
\end{bmatrix}
\]

\[
\text{var}(\nu_t) = Q = \sigma_e^2 \begin{bmatrix}
1 & 1 \\
1 & 1
\end{bmatrix}
\]

\[
y_t = H_t \xi_t + 0
\]

\[
H_t = \begin{bmatrix}
1 \\
0
\end{bmatrix} \text{ or } H_t = \begin{bmatrix}
0 & 0
\end{bmatrix}
\]

Example \( \sigma_e = 1, \phi = 0.9 \) and \( \theta = 0.5 \).
EX2.4: \( y(t), x(t) \) and \( x(t/T) \)
Example: **Nowcasting** (Good reference: Banbura, Giannoni, Modugno, and Reichlin (2013).)

- **Problem:** $y_t$ is a variable of interest (e.g., GDP growth rate in quarter $t$). It is available with a lag (say in $t+1$ or $t+2$). $X_t$ is a vector of variables that are measured *during* period $t$ (and perhaps earlier). How do you guess the value of $y_t$ given the $X$ data that has been revealed.

- ‘**Solution’**: Suppose $X_{t_1}$ denotes the information known at time $t_1$. Then best guess of $y_t$ is $\mathbb{E}(y_t | X_{t_1})$.
  - But how do you compute $\mathbb{E}(y_t | X_{t_1})$?
  - How do you update the estimate as another element of $X_t$ is revealed?
Giannone, Reichlin, et al modeling approach:

\[
\begin{bmatrix}
y_t \\
X_{1t} \\
\vdots \\
X_{nt}
\end{bmatrix} = \begin{bmatrix}
\lambda_y \\
\lambda_1 \\
\vdots \\
\lambda_n
\end{bmatrix} F_t + \begin{bmatrix}
e_{yt} \\
e_{1t} \\
\vdots \\
e_{nt}
\end{bmatrix}
\]

\[
\Phi(L)F_t = \eta_t
\]

- \( E(y_t| X_{t_1}) = \lambda_y \times E(F_t | X_{t_1}) \)
- \( E(F_t | X_{t_1}) \) computed by Kalman filter

(Lots of details left out)