Extending ARMA Models: Time Trends

Kevin Sheppard

Advanced Financial Econometrics: Forecasting

May 2020

Time Series Decomposition

- Interested in forecasting $X_{T+h|T}$
- Helpful to think about a decomposition

$$X_t = T_t + S_t + C_t + \epsilon_t$$

  - T_t is a deterministic time trend
  - S_t is a seasonal component
    - May be deterministic
  - C_t is a cyclic component
    - ARMA Component
    - May have seasonal lags
- Assume observed data is $\{X_1, \ldots, X_T\}$

Time Trends

- A basic trend model

$$X_t = \beta_0 + \beta_1 t + \epsilon_t$$

- This is a cross-sectional regression model
  - Time is just $1, 2, \ldots$
    - Makes no difference if you use a monotonic series with a constant difference
    - The actual year, $1990, 1991, 1992, \ldots$
    - Only affects the intercept
  - Might consider higher order trends

$$X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t$$

    - Rarely need higher order > 2
    - Higher order often indicates should use $\ln X_t$

Exponential Trends

- Models estimated in logs have exponential trends

$$\ln X_t = \beta_0 + \beta_1 t + \epsilon_t$$

- β_1 is the growth rate of X_t

$$X_t = \beta_0 \exp\left(\beta_1 t\right) \epsilon_t$$

- Pure trend models are simple to estimate using OLS

Forecasting

- Trend forecasting is simple

$$X_t = \beta_0 + \beta_1 t + \epsilon_t$$

- The forecast is then

$$\mathrm{E}_T\left[X_{T+h}\right] = \hat{X}_{T+h|T} = \beta_0 + \beta_1\left(T + h\right)$$

- We are often interested in prediction intervals
- A 95% Prediction interval should contain the truth 95% of the time
- Common to assume residuals are normally distributed

$$PI = \left[\hat{X}_{T+h|T} - 1.96\sigma, \hat{X}_{T+h|T} + 1.96\sigma\right]$$

- In pure time trend models the PI does not depend on h

Forecasting Exponential Trends

- Forecasts in exponential trend models is more involved
- **Assumption**: $\epsilon_t \overset{\text{i.i.d.}}{\sim} N\left(0, \sigma^2\right)$ so that the forecast variable is log-normal

Median

$$\hat{X}_{T+h|T} = \exp\left(\beta_0 + \beta_1 \left(T + h\right)\right)$$

Mean

$$\ln \hat{X}_{T+h|T} \sim N\left(\beta_0 + \beta_1 \left(T + h\right), \sigma^2\right) \Rightarrow \hat{X}_{T+h|T} \sim \text{LogNormal}\left(\beta_0 + \beta_1 \left(T + h\right), \sigma^2\right)$$

- Uses normality assumption of ϵ_t

$$\hat{X}_{T+h|T} = \exp\left(\beta_0 + \beta_1 \left(T + h\right) + \sigma^2/2\right)$$

  - $\exp\left(\cdot\right)$ is a convex function so Jensen's inequality applies

$$\mathrm{E}_T\left[\exp\left(\ln \hat{X}_{T+h|T}\right)\right] > \exp\left(\mathrm{E}_T\left[\ln \hat{X}_{T+h|T}\right]\right)$$

Prediction Intervals

- Prediction intervals are simple

$$PI = \left[\exp\left(\beta_0 + \beta_1\left(T+h\right) - 1.96\sigma\right), \exp\left(\beta_0 + \beta_1\left(T+h\right) + 1.96\sigma\right)\right]$$

- Symmetric in logs, asymmetric in levels
  - Quantiles are preserved under transformation
  - May not be possible to construct a symmetric PI that has a positive lower bound

Conclusions

- Trends are common in many time series
- Modeling the trend is essential when producing multi-step forecasts
- Trend estimation only requires OLS
- In practice trends should usually be limited to linear
  - Higher-order trends can produce large forecasting errors at longer horizons
- Key choice is whether to model the level of the log
- Forecasts of logged data can be produced using one of two methods
- Prediction intervals is simple in either case

Extending ARMA Models: Seasonality

Kevin Sheppard

Advanced Financial Econometrics: Forecasting

May 2020

Seasonality

- Pure seasonal model

$$X_t = S_t + \epsilon_t$$

- Seasonal pattern repeats every m observations
  - Traditionally defined on an annual basis
  - Can be defined over other frequencies
    - Day of Week (5 or 7)
    - Hour of Day
    - Week of Month
  - Common feature is that their occurence is completely predictable
- May have multiple seasonalities
  - Month, Day of Week, Hour of Day

Seasonal Dummies

- Basic deterministic seasonality uses dummy variables

$$X_t = \sum_{i=1}^{m} \gamma_i I_{[S_m(t)=i]} + \epsilon_t$$

  - $S_m(t) = t - m\lfloor (t-1)/m \rfloor$ which returns values in $1, \ldots, m$
- Alternative parameterization

$$X_t = \beta_0 + \sum_{i=1}^{m-1} \gamma_i I_{[S_m(t)=i]} + \epsilon_t$$

- Multiple Seasonalities use additive decomposition
- Assume seasonal frequencies of m_1 and m_2, $m_2 > m_1$, m_2 is not an integer multiple of m_1

$$X_t = \sum_{i=1}^{m_1} \gamma_i I_{[S_{m_1}(t)=i]} + \sum_{j=2}^{m_2} \delta_j I_{[S_{m_2}(t)=j]} + \epsilon_t$$

  - Must drop one dummy when using multiple seasons

Estimation

- Estimation is just OLS
- Simple to combine with time trends

$$X_t = \beta_1 t + \beta_2 t^2 + \sum_{i=1}^{m} \gamma_i I_{[S_m(t)=i]} + \epsilon_t$$

- Common to use an ANOVA-like test for seasonalities

$$\text{Restricted } X_t = \beta_0 + \epsilon_t$$

$$\text{Unrestricted } X_t = \beta_0 + \sum_{i=1}^{m-1} \gamma_i I_{[S_m(t)=i]} + \epsilon_t$$

- Null is $H_0 : \gamma_i = 0 \; i = 1, \ldots, m-1$
- Test using an F-test

$$\frac{R_U^2 - R_R^2}{1 - R_U^2} \times \frac{T-m}{m-1} \sim F_{[m-1,T-m]}$$

Forecasting and Prediction Intervals

- Forecasts are equally simple

$$\hat{X}_{T+h|T} = \beta_1 t + \beta_2 t^2 + \gamma_{S_m(T+h)}$$

- Predictions intervals are standard

$$PI = \left[\hat{X}_{T+h|T} - 1.96\sigma, \hat{X}_{T+h|T} + 1.96\sigma\right]$$

  and do not depend on h
- If modeling $\ln X_t$ can use the mean or median forecast

Fourier Series

- Fourier Series are an alternative to dummy variables
- Provide smooth seasonal effects unlike dummies
- Particularly useful when the season has many periods
  - Weekly seasonality in a year
  - Hourly seasonality in a week
- Choose order of Fourier, K

$$X_t = \sum_{k=1}^{K} \gamma_k \cos\left(2k\pi \frac{S_m(t)}{m}\right) + \delta_k \sin\left(2k\pi \frac{S_m(t)}{m}\right) + \epsilon_t$$

  - In practice, K is small
  - Choose using information criterion
  - Only fully general when $K = m/2$
- Simple to combine more than one seasonality using $m_1, m_2, \ldots$
- Forecast replaces t with $T + h$

$$\hat{X}_{T+h|t} = \sum_{k=1}^{K} \gamma_k \cos\left(2k\pi \frac{S_m(T+h)}{m}\right) + \delta_k \sin\left(2k\pi \frac{S_m(T+h)}{m}\right)$$

Conclusions

- Seasonal dummies account for seasonal shifts in a time series
- Easy to build a model with seasonal dummies and time trends
- Seasonal dummies do not affect prediction intervals
- Fourier seasonal allow for parsimonious specification of seasonality
  - Important when the period of a series is large
- Multiple seasonalities can be captured using combinations of the two approaches

Extending ARMA Models: Seasonal Autoregressions

Kevin Sheppard

Advanced Financial Econometrics: Forecasting

May 2020

The Lag Operator

- The *Lag Operator*, L, is essential to understanding Seasonal ARMAs
- Key properties

$$
\begin{aligned}
LX_t &= X_{t-1} \\
L^2 X_t &= L\left(LX_t\right) = LX_{t-1} = X_{t-2} \\
L^p L^q X_t &= L^{p+q} X_t = X_{t-(p+q)}
\end{aligned}
$$

- Familiar models written with *Lag Polynomials*
  - AR(1)

$$
\begin{aligned}
X_t &= \phi_0 + \phi_1 X_{t-1} + \epsilon_t \\
X_t - \phi_1 X_{t-1} &= \phi_0 + \epsilon_t \\
X_t - \phi_1 L X_t &= \phi_0 + \epsilon_t \\
(1 - \phi L)\, X_t &= \phi_0 + \epsilon_t
\end{aligned}
$$

  - AR(P)

$$
\left(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_P L^p\right) X_t = \phi_0 + \epsilon_t
$$

$$
X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-P} + \epsilon_t
$$

Seasonal AR Models

- Pure Seasonal AR

$$(1 - \phi_m L^m) X_t = \phi_0 + \epsilon_t$$
$$X_t = \phi_0 + \phi_m X_{t-m} + \epsilon_t$$

  - This model is not plausible
  - Equivalent to m unrelated AR(1) models iterwoven

- Seasonal AR with short-run dynamics

$$(1 - \phi_1 L)(1 - \phi_m L^m) X_t = \phi_0 + \epsilon_t$$
$$\left(1 - \phi_1 L - \phi_m L^m + \phi_1 \phi_m L^{(m+1)}\right) X_t = \phi_0 + \epsilon_t$$

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_m X_{t-m} - \phi_1 \phi_m X_{t-m-1} + \epsilon_t$$

  - Restricted AR$(m+1)$
  - Sometimes written as a SAR$(1) \times (1)$
  - Generally SAR$(P) \times (P_S)$

Ignoring the Restriction

- Can always estimate unrestricted model

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_m X_{t-m} + \phi_{m+1} X_{t-m-1} + \epsilon_t$$

  - Could even test $H_0 : \phi_{m+1} = -\phi_1 \phi_m$
    - Not important in forecasting
  - An information criterion can be used to select between these two
  - Forecasting is standard for AR models using standard representation
- Unrestricted model can be estimated using OLS
- Restricted model requires a constrained estimator, e.g., NLLS

Prediction Intervals

- Prediction intervals are **not** constant

$$PI = \left[X_{T+h|T} \pm 1.96\tilde{\sigma}_h\right]$$

  - $\tilde{\sigma}_h$ is a function of horizon and model parameters
  - Simple to compute using MA(∞) representation
- Recall *companion form* of AR(P)

$$\begin{bmatrix} X_t - \mu \\ X_{t-1} - \mu \\ X_{t-2} - \mu \\ \vdots \\ X_{t-P+1} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \ldots & \phi_P \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} \begin{bmatrix} X_{t-1} - \mu \\ X_{t-2} - \mu \\ X_{t-3} - \mu \\ \vdots \\ X_{t-P} - \mu \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

  - $\mu = \phi_0 / \left(1 - \phi_1 - \phi_2 - \ldots - \phi_P\right)$

The MA(∞) Representation

$$\mathbf{Z}_t = \mathbf{\Phi}\mathbf{Z}_{t-1} + \boldsymbol{\eta}_t$$

- The MA(∞) representation is then

$$\begin{aligned}\mathbf{Z}_t &= \boldsymbol{\eta}_t + \mathbf{\Phi}\boldsymbol{\eta}_{t-1} + \mathbf{\Phi}^2\boldsymbol{\eta}_{t-2} + \mathbf{\Phi}\boldsymbol{\eta}_{t-3} + \ldots \\ &= \mathbf{\Xi}_0\boldsymbol{\eta}_t + \mathbf{\Xi}_1\boldsymbol{\eta}_{t-1} + \mathbf{\Xi}_2\boldsymbol{\eta}_{t-2} + \mathbf{\Xi}_3\boldsymbol{\eta}_{t-3} + \ldots\end{aligned}$$

  - Define $\xi_j = \mathbf{\Xi}_j^{[1,1]}$ as the $(1,1)$ element, then

$$\tilde{\sigma}_h^2 = \sigma^2\left(1 + \xi_1^2 + \xi_2^2 + \ldots + \xi_{h-1}^2\right)$$

- Easy to show in the AR(1)

$$X_t = \phi_0 + \phi_1 X_{t-1} + \epsilon_t$$

$$\tilde{\sigma}_h^2 = \sigma^2\left(1 + \phi^2 + \phi^4 + \ldots + \phi^{2(h-1)}\right)$$

- General formula for impulses for AR(P) in VAR slides and notes

Random Walks with Seasonality

- A seasonal random walk has a unit root at the seasonal frequency

$$(1 - L^m) X_t = \epsilon_t$$

- Need short run-dynamics to make plausible

$$(1 - \phi_1 L) \left(1 - L^M\right) X_t = \epsilon_t$$

- Seasonal Unit Roots need seasonal differencing

$$\Delta_m X_t = X_t - X_{t-m}$$

  - Note that Δ^m and Δ_m are different

$$\begin{aligned}
\Delta^m X_t &= \Delta \left(\Delta^{m-1}\right) X_t = \Delta \left(\Delta \left(\Delta \left(\ldots \left(\Delta X_t\right)\right)\right)\right) \\
\Delta_m X_t &= X_t - X_{t-m} \\
\Delta^2 X_t &= \Delta \left(X_t - X_{t-1}\right) = X_t - 2X_{t-1} + X_{t-2} \\
\Delta_2 X_t &= X_t - X_{t-2}
\end{aligned}$$

Seasonal Differencing

- Seasonal difference removes seasonal unit roots

$$\Delta_m X_t = X_t - X_{t-m} = (1 - L^m)\, X_t$$

- so that

$$\begin{aligned}(1 - \phi_1 L)\, \Delta_m X_t &= \epsilon_t \\ \Delta_m X_t &= \phi_1 \Delta_m X_{t-1} + \epsilon_t \\ \tilde{X}_t &= \phi_1 \tilde{X}_{t-1} + \epsilon_t\end{aligned}$$

- **Note**: When you use seasonal differences, you do not need seasonal dummies

Conclusions

- Seasonal Autoregressions (SAR) capture dynamics at the seasonal frequency
- Combined with short run dynamics to construct plausible models
- Unrestricted models which include the same terms are simple to estimate using OLS
- Prediction intervals depend on the parameters of the MA(∞) representation
  - These are the impulses
- Seasonal random walks are removed using seasonal differencing
- Seasonally differencing also removes level shifts series
  - No need to use both seasonal dummies and differencing

Extending ARMA Models: Seasonal Moving Averages

Kevin Sheppard

Advanced Financial Econometrics: Forecasting

May 2020

Seasonal MA

- Seasonality can be introduced into MA using the same structure
- Seasonal MA$(1) \times (1)$

$$
\begin{aligned}
X_t &= (1 + \theta_1 L)\,(1 + \theta_m L^m)\,\epsilon_t \\
&= \left(1 + \theta_1 L + \theta_m L^m + \theta_1 \theta_m L^{m+1}\right)\epsilon_t \\
&= \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_m \epsilon_{t-m} + \theta_1 \theta_m \epsilon_{t-m-1}
\end{aligned}
$$

- Restricted MA$(m+1)$
- Less common in forecasting since unrestricted SAR can be estimated using OLS

Prediction Intervals in MA Models

- Prediction intervals for MA processes are simple

$$X_t = \mu + \sum_{i=1}^{Q} \theta_i \epsilon_{t-i} + \epsilon_t$$

- The h-step error is then

$$X_{T+h} - \hat{X}_{T+h|T} = \epsilon_{T+h} + \sum_{i=1}^{\min(h-1,Q)} \theta_i \epsilon_{T+h-i}$$

- The variance of the forecast error

$$\sigma_h^2 = \sigma^2 \left(1 + \sum_{i=1}^{\min(h-1,Q)} \theta_i^2 \right)$$

- Prediction intervals are then

$$\left[\hat{X}_{T+h|T} \pm 1.96\sigma_h \right]$$

MA Invertibility and Prediction

- Inverting MAs help understand MA prediction
- We only observe $\{X_t\}$

$$\begin{aligned}
X_t &= \theta_1 \epsilon_{t-1} + \epsilon_t \\
X_{t-1} &= \theta_1 \epsilon_{t-2} + \epsilon_{t-1} \Rightarrow \epsilon_{t-1} = X_{t-1} - \theta_1 \epsilon_{t-2} \\
X_t &= \epsilon_t + \theta_1 \left(X_{t-1} - \theta_1 \epsilon_{t-2} \right) \\
&= \epsilon_t + \theta_1 X_{t-1} - \theta_1^2 \epsilon_{t-2} \\
\epsilon_{t-2} &= X_{t-2} - \theta_1 \epsilon_{t-3} \\
X_t &= \epsilon_t + \theta_1 X_{t-1} - \theta_1^2 \left(X_{t-2} - \theta_1 \epsilon_{t-3} \right) \\
&= \epsilon_t + \theta_1 X_{t-1} - \theta_1^2 X_{t-2} + \theta_1^3 \epsilon_{t-3}
\end{aligned}$$

- Continuing back to $t = 1$,

$$X_t = \epsilon_t + \sum_{i=1}^{t-1} (-1)^{i+1} \theta^i X_{t-i} + (-1)^{t+1} \theta^t \epsilon_0$$

- Assuming $\epsilon_0 = 0$, this is an AR(t)

Forecasts from MA models

- The optimal one-step forecast is then

$$\hat{X}_{T+1|T} = \theta X_T - \theta^2 X_{T-1} + \theta^3 X_{T-3} + \ldots + (-1)^{T-1} \theta^T X_1$$

  - Only depends on observed values
- Longer-horizon prediction recursively applies this AR(T)
- If mean is not 0:
  - Subtract μ from X_t
  - Produce optimal forecast $\hat{\hat{X}}_{T+h|T}$ for $\tilde{X}_t = X_t - \mu$
  - Add mean back $\hat{X}_{T+h|T} = \mu + \hat{\hat{X}}_{T+h|T}$

Conclusions

- Optimal forecasts in MA models only depend on observed data
- The Seasonal MA adds seasonal lags like a Seasonal Autoregression
- Prediction intervals in MA models are simple functions of the MA parameters

Extending ARMA Models: SARIMA
Seasonal Autoregressice Integrated Moving Averages

Kevin Sheppard

Advanced Financial Econometrics: Forecasting

May 2020

Seasonal ARMA Models

- Can apply seasonalities to both components in an ARMA
- Seasonal ARMA$(P, Q) \times (P_s, Q_s)$
- Seasonal ARMA$(1, 1) \times (1, 1)$

$$(1 - \phi_1 L)\left(1 - \phi_m L^m\right) X_t = (1 + \theta_1 L)\left(1 + \theta_m L^m\right)\epsilon_t$$

Incorporating the differencing parameter

- Common to also incorporate *differencing* order D into specification
- **S**easonal **A**utoregression **I**ntegrated **M**oving **A**verage (**SARIMA**)
- Each order has three parameters

$$(P, D, Q) \times (P_s, D_s, Q_s)$$

- In practice one of D or D_s is usually 0, other is either 1 or 0
  - The D and D_s parameters indicate how to difference
  - D uses the standard difference operator
  - D_s applies the seasonal differecne operator
- If X_t is SARIMA$(P, 1, Q) \times (P_s, 0, Q_s)$, then ΔX_t is SARIMA$(P, 0, Q) \times (P_s, 0, Q_s)$
- If X_t is SARIMA$(P, 0, Q) \times (P_s, 1, Q_s)$, then $\Delta_m X_t$ is SARIMA$(P, 0, Q) \times (P_s, 0, Q_s)$

Forecasting

- Order of integration matters for forecasting and prediction intervals
- For non-seasonal differenced series

$$\hat{X}_{T+h|T} = X_T + \sum_{i=1}^{h} \mathrm{E}_T \left[\Delta X_{T+i}\right]$$

- Prediction Intervals have the form

$$PI = \left[\hat{X}_{T+h|T} \pm 1.96\breve{\sigma}_h\right]$$

$$\breve{\sigma}_h^2 = \sigma^2 \sum_{i=1}^{h} \left(1 + \sum_{j=1}^{\mathbf{i-1}} \xi_j\right)^2$$

- In a model with order $(1,1,0) \times (0,0,0)$ this is

$$\breve{\sigma}_h^2 = \sigma^2 \left\{(1)^2 + (1+\phi_1)^2 + \ldots + \left(1 + \phi_1 + \phi_1^2 + \ldots + \phi_1^{h-1}\right)^2\right\}$$

- Prediction intervals will continue to widen as the horizon increases
  - Reflects the unit root (random walk) in the time series

Forecasting with Seasonal Differencing

- In a Seasonally Differenced model we model $\Delta_m X_t$ so that

$$\mathrm{E}_T\left[X_{T+1}\right] = X_{T+1-m} + \underbrace{\mathrm{E}_T\left[\Delta_m X_{T+1}\right]}_{\text{1-step from model}}$$

$$= X_{T+1-m} + \mathrm{E}_T\left[X_{T+1}\right] - \mathrm{E}_T\left[X_{T+1-m}\right]$$

$$= \mathrm{E}_T\left[X_{T+1|T}\right] + \underbrace{X_{T+1-m} - X_{T+1-m|T}}_{0}$$

- In general

$$\hat{X}_{T+h|T} = X_{T+1-m} + \sum_{i=1}^{h} \mathrm{E}_T\left[\Delta_m X_{T+h|T}\right]$$

- Note that $\Delta_m X_t$ is the LHS in the seasonally differenced model

The Complete Model

- Start by transforming X_t
  - Log or level
  - Level, Difference, or Seasonal Difference

$$
\begin{aligned}
Y_t = \quad & \text{Constant} \\
& + \text{Trend} \\
& + \text{Seasonal Dummies} \\
& + \text{AR} + \text{Seasonal AR} \\
& + \text{MA} + \text{Seasonal MA} + \epsilon_t
\end{aligned}
$$

- Recommendations for forecasting
  - Differencing, Trends, and Seasonal Dummies are essential for multi-step forecasting
  - ARMA terms matter for shorter horizons
  - Always difference if “close” to a unitroot

Conclusions

- SARIMA is a unified framework for modeling trends, seasonal and cyclical components
- Differencing, trend and seasonal specification are keys to good forecasting models
  - Especially true over longer horizons
- Forecasts from models built using differenced data accumulate the forecast differences
- Prediction intervals also depend on sums of accumulated MA(∞) parameters