# Estimation and Hypothesis Testing with Unfiltered and Filtered Data

## Ghysels & Osborn Chapter 5

Nora Prean

Econometrics of seasonality, May 2012

# Adjusted versus unadjusted data

- Seasonally adjusted data
  - e.g. structural models designed to capture business cycle fluctuations
- Unadjusted data
  - models should reflect decision-making process of representative agents **facing seasonal fluctuations**

- Question will remain unsolved; BUT researchers should be aware of econometric consequences of using seasonally unadjusted and seasonally adjusted data

# Outlook

1. Misspecification of stochastic structure of seasonality in linear regression models
2. Consequences of using seasonally adjusted (filtered) data
3. Conclusion

# Sources of misspecification

Misspecifying seasonality as deterministic when in fact it is stochastic and nonstationary and vice versa

Assume

$$y_t = \sum_{s_1}^{S} \delta_{st}\gamma_s + z_t \qquad (5.1)$$

where $\phi(L)z_t = \theta(L)\epsilon_t, \quad \epsilon_t \sim \text{iid}(0, \sigma^2)$

- Inappropriate seasonal differencing:
  - if seasonality is deterministic (all roots outside unit circle): run regression in (5.1)
  - corresponding seasonal difference model:
    $\Delta_S y_t = \Delta_S z_t$ (5.2)
  - seasonal differences operator $\Delta_S$ removes seasonal means (no longer identified)
  - furthermore, the $\Delta_S$ filter results in $\Delta_S z_t$ ("overdifferencing" of the data, see Maravall, 1995)

# Misspecification, cont.

- Inappropriate seasonal dummies [$\phi(L)$ has roots on unit circle at seasonal frequencies]:
  - "spurious deterministic seasonality" (Abeysinghe, 1991, 1994; Franses, Hylleberg, Lee (1995)
  - R2 misleading (estimation of (5.1) might lead to high R2 even if DGP is nonstationary seasonal process)
  - conventional F test for the null $\gamma_1 = ... = \gamma_S$ not valid

# Excursus: filtering

- Idea: filter raw data such that seasonal fluctuations disappear from the series
- Long tradition of decomposing time series in mutually orthogonal unobserved components
  - any seasonal adjustment procedure rests on specific decomposition of a series into a **"trend cycle"**, and **seasonal** and **irregular components**
  - multiplicative, pseudo-additive, (log) additive decomposition
- Linear X-11 filter (U.S. Census Bureau, N.B.E.R), for monthly and quarterly data
  - 3 stage procedure, each with several sub-stages, using different MA filters, seasonal moving average filters,...

# Regression models and Filtering

Consider:

$$\begin{aligned} y_t &= \textstyle\sum_{j=1}^{p} \alpha_j y_{t-j} + x_t' \beta + \delta_t + \epsilon \\ t &= -l, ..., 0, 1, ...T + k \end{aligned} \tag{5.6}$$

regressors non-stochastic, error process $\epsilon' = (\epsilon_{-l}, ..., \epsilon_{T+k})$ has known covariance matrix $E(\epsilon\epsilon') = \Omega$

Assume that data is filtered with a known and linear filter; filter weights $(v_{-l}, ..., v_k)$

# Filter matrix $F$

$(T+1) \times (T+k+l+1)$ filter matrix $F$:

$$F = \begin{pmatrix} v_{-l} & \dots & \dots v_k & & 0 \\ & \ddots & & \ddots & \\ & & \ddots & & \ddots \\ 0 & & v_{-l} & & v_k \end{pmatrix}$$

- Filter matrix $F$ transforms sample of size $\mathcal{T} = T + k + l + 1$ into filtered data set with $T + 1$ observations (at each end of sample data are discarded by the two-sided filter)
- Symmetric, two-sided filter ($v_{-l} = v_l$); filtering weights sum to one ($\sum_{i=-m}^{m} v_i = 1$ with $m = k = l$)
- Constant and time trend in regression model invariant to filtering ($X_F = X$)

- Define filtered and unfiltered data sets as

$$\begin{aligned} y' &= (y_{-l}, ..., y_{T+k}), \mathcal{T} \times 1, \\ y_F &= Fy, (T+1) \times 1, \text{and} \\ y_U &= Uy, (T+1) \times 1; \\ U &\equiv (O_{(T+1)\times l} I_{T+1} O_{(T+1)\times k}) \end{aligned}$$

- Matrix $U$ cuts away observations at each end of the sample so that $y_F$ and $y_U$ are filtered and unfiltered data sets of equal sizes $T+1$ with disturbances drawn from the same random vector $\epsilon$

# Simple case: no lagged dependent variables

- Suppose $y_t = x_t'\beta + \epsilon_t, \qquad t = -l, ..., 0, 1, ..., T + k$
- $\Omega$ and $F$ known, thus we know

$$\begin{aligned} E(\epsilon_U \epsilon_U') &\equiv \Omega_U = & U\Omega U \\ E(\epsilon_F \epsilon_F') &\equiv \Omega_F = & F\Omega F \end{aligned}$$

- $\hat{\beta}_j = (X_j'X_j)^{-1}X_j'y_j, \qquad j = U, F$
- $\hat{\beta}$ unbiased: $E(\hat{\beta}_F) = (X_F'X_F)^{-1}X_F'E(X_F\beta + F\epsilon) = \beta$
- and $Var(\hat{\beta}_F) = (X_F'X_F)^{-1}X_F'\Omega_F X_F(X_F'X_F)^{-1}$

# Filtering ARIMA types of models

- Suppose

$$y_t = \mu + \gamma_t + z_t$$

and linear filter $v(L)$ where $v(L) = v(L^{-1})$ and $v(1) = 1$

- Filtered version of $y_t$ an be written as:

$$\begin{array}{rcll} y_t & = & z_t, & y_0 = 0 \\ y_t^F & = & v(L)z_t, & y_0^F = 0 \end{array}$$

i.e. $y_t^F = v(L)y_t$

- Then (consider simple case) $y_t = \alpha y_{t-1} + \epsilon_t$
- $\longrightarrow$ compare $\hat{\alpha}$ with $\hat{\alpha}_F$ to determine asymptotic bias

- Filtering of unit root processes:
  - when $\phi(L)$ has root on unit circle, $\hat{\alpha}$ and $\hat{\alpha}_F$ both converge to one
  - thus, NO bias $\longrightarrow$ asymptotic null distributions of unit root test statistics not affected by filtering of the data
  - HOWEVER, finite sample distributions are affected by filtering

- Filtering Stationary ARMA processes:
  - when $\phi(L)$ does not contain unit root $\longrightarrow$ asymptotic bias induced by filtering
  - Ghysels and Perron (1993) show: positive bias in case of linear X-11 filter (thus, unit root tests performed with filtered data can be expected to be less powerful against stationary alternatives)

# Conclusions

- Use of adjusted (filtered) data
  - discards important information about economic dynamics
  - possibly introduces severe biases
- Use of unfiltered data
  - incorporates misspecification risk
  - standard statistical inference procedures often result in close fits at seasonal frequencies but underemphasizes other frequencies (Sims, 1974, 1993); loss function might put a heavy penalty on misfitting the seasonal frequencies
  - Hansen and Sargent (1993) exploit further sources of misspecification; their findings seem to support Sims (1974, 1993)
  - however, debate still ongoing; depends on the context