Why You Should Never Use the Hodrick-Prescott Filter

James D. Hamilton

University of California, San Diego

Here's why

1. HP introduces spurious correlations that have no basis in the true data-generating process.
2. Filtered values at the end of the sample are very different from those in the middle, and are also characterized by spurious dynamics.
3. A statistical formalization produces estimates of smoothing parameter λ around 1 (not 1600).
4. There's a better alternative.

Outline

Characterization A. HP fits smooth trend to data

Choose trend g_t that is as close as possible to the observed series y_t with a penalty for changing the trend too quickly.

$$\min_{\{g_t\}_{t=-1}^{T}} \left\{ \sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=1}^{T} [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \right\}.$$

Common practice: $\lambda = 1600$ for quarterly data.

Solution: $g_t = a_t^{*\prime} y$ for $y = (y_1, y_2, ..., y_T)'$.

B. Alternative characterization: make statistical assumptions about trend and cycle and find optimal estimate given data

$$\min_{a_t} E(g_t - a_t' y)^2$$

$$\Delta^2 g_t = v_t$$

$$y_t - g_t = c_t$$

Proposition: if assume that v_t and c_t are uncorrelated white noise, then HP trend and optimal statistical estimate of trend are numerically identical when $\sigma_c^2/\sigma_v^2 = \lambda$.

In other words, can calculate HP filter using Kalman smoother for this assumed model.

However, if data were generated by this process, inferred cycle c_t would be white noise (no discernible patterns).

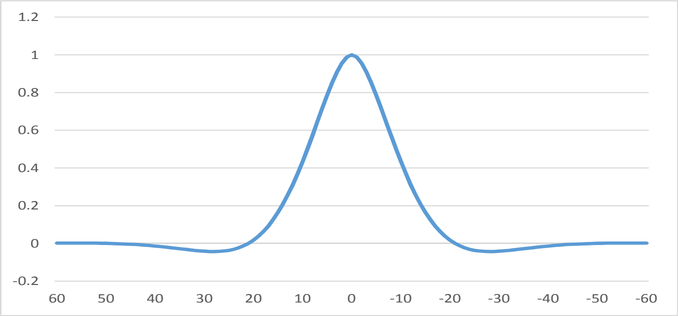
C. Algebraic representation

For t more than about 60 observations away from 1 or T, can calculate HP cyclical component $c_t = y_t - g_t^*$ as

$$c_t = C_0 \left[\frac{1 - (\phi_1^2/4)L}{1 - \phi_1 L - \phi_2 L^2} + \frac{1 - (\phi_1^2/4)L^{-1}}{1 - \phi_1 L^{-1} - \phi_2 L^{-2}} - 1 \right] \Delta^4 y_{t+2}$$
$$= C_0 \left[-\Delta^4 y_{t+2} + \sum\nolimits_{j=0}^{\infty} R^j [\cos(mj) + \cot(m) \sin(mj)] q_t^{(j)} \right]$$

$$q_t^{(j)} = \Delta^4 y_{t+2-j} - (\phi_1^2/4)\Delta^4 y_{t+1-j} + \Delta^4 y_{t+2+j} - (\phi_1^2/4)\Delta^4 y_{t+3+j}$$

$0 < \phi_1 < 2$, $-1 < \phi_2 < 0$, $0 < R < 1$


1.2
1
0.8
0.6
0.4
0.2
0
-0.2
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60

2. Consequences: what kind of time-series processes are we likely to encounter in economics?

Simple economic theory suggests lots of variables should behave like random walk:

- Stock prices (Fama, 1965)
- Futures prices (Samuelson, 1965)
- Long-term interest rates (Sargent, 1976; Pesando, 1979)
- Oil prices (Hamilton, 2009)
- Consumption spending (Hall, 1978)
- Inflation, tax rates, and money supply growth rates (Mankiw, 1987)

A random walk is often hard to beat in out-of-sample forecasting exercises

- Exchange rates (Meese and Rogoff, 1983; Cheung, Chinn and Pascual, 2005)
- Stock prices (Flood and Rose, 2010)
- Inflation (Atkeson and Ohanian, 2001)
- GDP (Balcilar, et al., 2015)

Conclusion: If HP is not a good approach for a random walk, we shouldn't be using it on economic time series

Consequences of applying HP to random walk (Cogley and Nason, 1995)

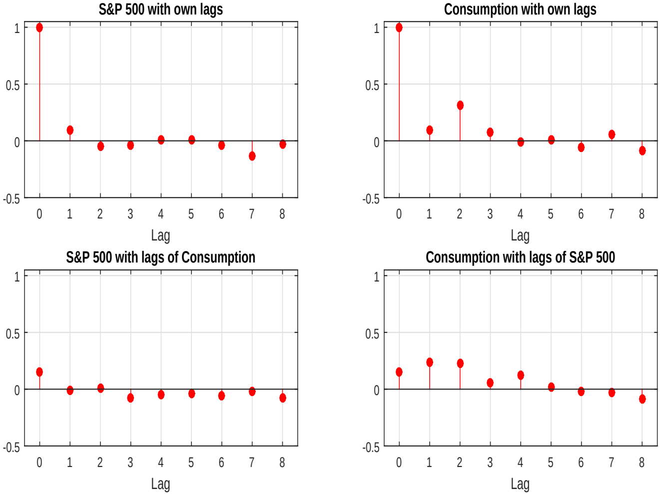
HP:

$$c_t = C_0 \left[\frac{1 - (\phi_1^2/4)L}{1 - \phi_1 L - \phi_2 L^2} + \frac{1 - (\phi_1^2/4)L^{-1}}{1 - \phi_1 L^{-1} - \phi_2 L^{-2}} - 1 \right] \Delta^4 y_{t+2}$$

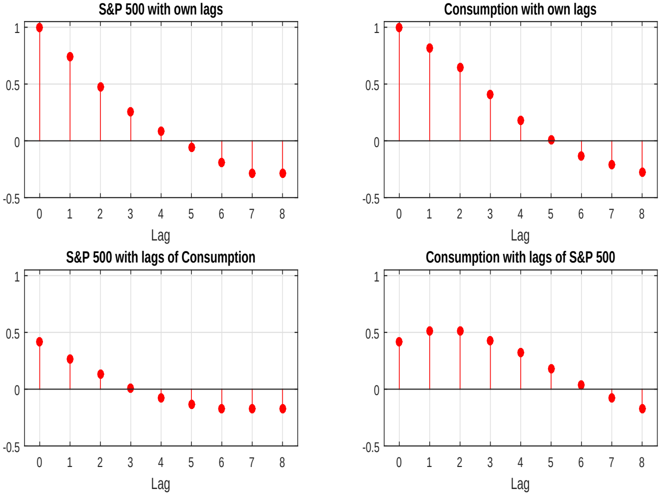
Random walk: $\Delta y_t = \varepsilon_t$ is completely unpredictable

$$c_t = C_0 \left[\frac{1 - (\phi_1^2/4)L}{1 - \phi_1 L - \phi_2 L^2} + \frac{1 - (\phi_1^2/4)L^{-1}}{1 - \phi_1 L^{-1} - \phi_2 L^{-2}} - 1 \right] \Delta^3 \epsilon_{t+2}$$

Example: autocorrelations and cross-correlations of Δy_t for $y =$ log stock price and log real consumption



Autocorrelations and cross-correlations of HP cyclical component



3. A one-sided HP filter

- HP g_t and c_t can "predict" the future because they are a function of the future.
- Could use the Kalman filter instead of the Kalman smoother to implement a one-sided version of HP.

Kalman smoother:

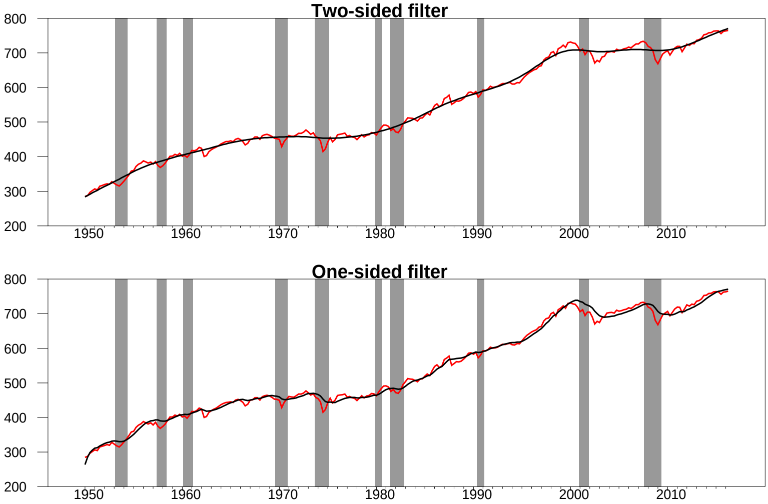
$$\min_{a_{t1},\ldots a_{tT}} E(g_t - a_{t1}y_1 - \cdots - a_{tT}y_T)^2$$

Kalman filter:

$$\min_{a_{t1},\ldots a_{tt}} E(g_t - a_{t1}y_1 - \cdots - a_{tt}y_t)^2$$

Kalman filter numerically identical to HP trend for date t using data $(y_1, \ldots, y_t)$.

Example: log of S&P 500



4. Estimating smoothing parameter

If assume that $\Delta^2 g_t$ and c_t are white noise, could use Kalman filter to find quasi-maximum likelihood estimates of σ_c^2 and σ_v^2.

	σ_c^2	σ_v^2	λ
GDP	0.115	0.468	0.245
Consumption	0.163	0.174	0.940
Investment	4.187	12.196	0.343
Exports	5.818	3.341	1.741
Imports	4.423	4.769	0.927
Government spending	0.221	1.160	0.191
Employment	0.006	0.250	0.023
Unemployment rate	0.014	0.092	0.152
GDP Deflator	0.018	0.081	0.216
S&P 500	21.284	15.186	1.402
10-year Treasury yield	0.135	0.054	2.486
Fed Funds Rate	0.633	0.116	5.458
Real Rate	0.875	0.091	9.596

Conclusion

- HP is very inappropriate for a random walk.
- As commonly applied, it is not even appropriate for the only example for which anyone has claimed it should work well!

5. A better alternative

- Proposed definition of trend:
  - Component that could be predicted 2 years earlier
- Proposed definition of cycle:
  - Error associated with a 2-year-ahead forecast
- Justification:
  - The primary reason well be wrong in a 2-year-ahead forecast is timing of recession or recovery

- But don't we need a model of the trend to make a 2-year-ahead forecast?
- Answer: no!

Proposition: if $\Delta^d y_t$ is stationary for some d, then can write y_{t+h} as a linear function of $y_t, y_{t-1}, ..., y_{t-d+1}$ plus a stationary residual.

Example: $d = 1$

$$u_t = \Delta y_t \sim I(0)$$

$$y_{t+h} = y_t + u_{t+1} + u_{t+2} + \cdots + u_{t+h} = y_t + w_t^{(h)}$$

$$w_t^{(h)} = u_{t+1} + u_{t+2} + \cdots + u_{t+h} \sim I(0)$$

Example: $d = 2$

$$u_t = \Delta^2 y_t \sim I(0)$$

$$y_{t+h} = (h+1)y_t - hy_{t-1} + u_{t+h} + 2u_{t+h-1} + \cdots + hu_{t+1}$$

$$= (h+1)y_t - hy_{t-1} + w_t^{(h)}$$

$$w_t^{(h)} = u_{t+h} + 2u_{t+h-1} + \cdots + hu_{t+1} \sim I(0)$$

Easy to estimate the coefficients

If $y_t \sim I(2)$, what happens if we regress y_{t+h} on $(1, y_t, y_{t-1})'$?

- If coefficient on $y_t = h + 1$ and coefficient on $y_{t-1} = -h$, then average squared residual will tend to a finite number.
- For any other coefficients, average squared residual will tend to an infinite number.
- OLS will give a consistent estimate of parameters that characterize the trend.

We don't need to know d

If $y_t \sim I(2)$, what happens if we regress y_{t+h} on $(1, y_t, y_{t-1}, y_{t-2}, y_{t-3})'$?

- Two of the coefficients will make the residuals stationary.
- Other two coefficients will then try to forecast stationary component.

Conclusion: we don't need to know d.

Summary

If $y_t \sim I(d)$ for some unknown $d \leq 4$, the population linear projection of y_{t+h} on $(1, y_t, y_{t-1}, y_{t-2}, y_{t-3})'$ exists and can be consistently estimated by OLS regression.

Proposal

For quarterly data estimate by OLS $y_{t+8} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \beta_4 y_{t-3} + v_{t+8}$ and interpret the residuals $\hat{v}_{t+8}$ as the cyclical component.

Example 1: Random walk

Population values would be $\beta_1 = 1$ and all other $\beta_j = 0$.

$$v_{t+8} = y_{t+8} - y_t$$

Note this filter also eliminates seasonal components.

Example 2: What if we did this regression on data generated from a stationary DSGE?

If effects of shocks in theoretical model die out after 2 years, in data generated by theoretical model, coefficients in regression of y_{t+8} on $y_t, y_{t-1}, y_{t-2}, y_{t-3}$ will be zero and residuals will be deviations from steady state.

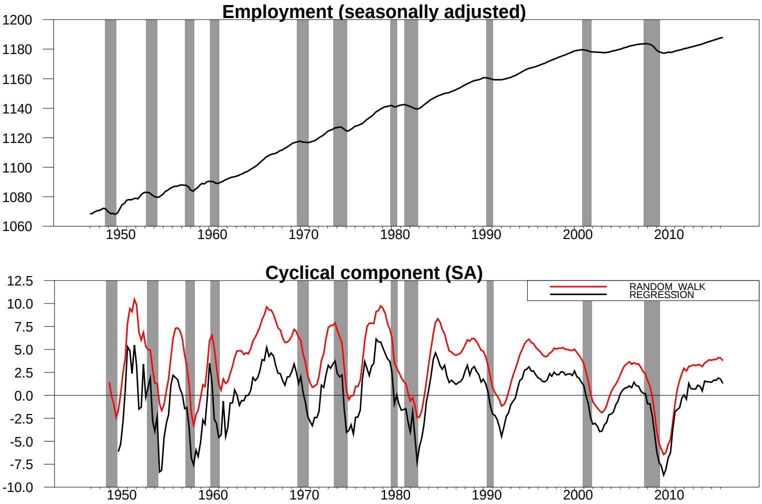
Can calculate analogous magnitude with observed nonstationary data.

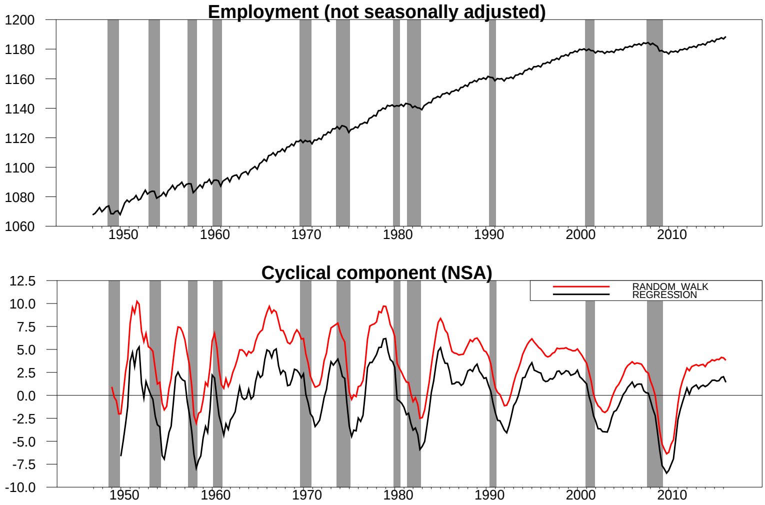
Example 3: Deterministic time trend

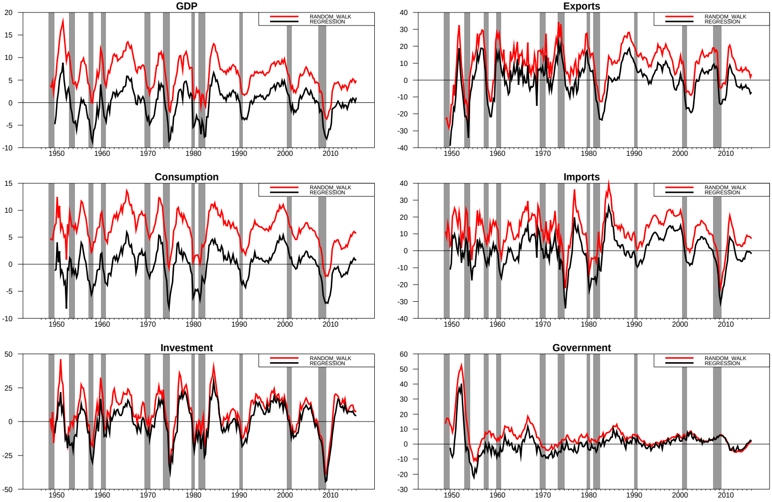
$y_t = \delta_0 + \delta_1 t + \varepsilon_t$ for ε_t white noise

Coefficients on $y_t, \ldots, y_{t-p+1}$ converge to $1/p$ and the implied trend for y_{t+h} is

$$\delta_0 + \delta_1(t+h) + p^{-1}(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_{t-p+1}).$$


Employment (seasonally adjusted)
1200
1180
1160
1140
1120
1100
1080
1060
1950
1960
1970
1980
1990
2000
2010
Cyclical component (SA)
RANDOM_WALK
REGRESSION
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-10.0


Employment (not seasonally adjusted)
1200
1180
1160
1140
1120
1100
1080
1060
1950
1960
1970
1980
1990
2000
2010
Cyclical component (NSA)
RANDOM_WALK
REGRESSION
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-10.0


GDP
Exports
Consumption
Imports
Investment
Government
RANDOM_WALK
REGRESSION
1950
1960
1970
1980
1990
2000
2010

Standard deviation of cyclical component and correlation with cyclical component of GDP

	Regression Residuals		Random walk	
	St. Dev.	GDP Corr.	St. Dev.	GDP Corr.
GDP	3.38	1.00	3.69	1.00
Consumption	2.85	0.79	3.04	0.82
Investment	13.19	0.84	13.74	0.80
Exports	10.77	0.33	11.33	0.30
Imports	9.79	0.77	9.98	0.75
Government spending	7.13	0.31	8.60	0.38
Employment	3.09	0.85	3.32	0.85
Unemployment rate	1.44	-0.81	1.72	-0.79
GDP Deflator	2.99	0.04	4.11	-0.13
S&P 500	21.80	0.41	22.08	0.38
10-year Treasury yield	1.46	-0.05	1.51	0.08
Fed funds rate	2.78	0.33	3.03	0.40
Real rate	2.25	0.39	2.60	0.42

6. Conclusion

HP tries to construct a stationary component from an I(4) series, but at a great cost.

- Introduces spurious dynamic relations that are purely an artifact of the filter.
- There exists no plausible data-generating process for which common popular practice would provide an optimal decomposition into trend and cycle.

There is an alternative approach that can also isolate a stationary component from an I(4) series.

- Preserves the underlying dynamic relations.
- Consistently estimates well defined population characteristics for a broad class of possible data-generating processes.