PRMED • The World Bank • 1818 H St NW • Washington DC 20433

The Hodrick-Prescott Filter

TREND-CYCLE DECOMPOSITIONS are routine in modern macroeconomics. The basic idea is to decompose the economic series of interest (*e.g.*, the log of GDP) into the sum of a slowly-evolving secular trend and a transitory deviation from it which is classified as 'cycle':

$$\begin{array}{ccccc} x_t & = & \tau_t & + & \zeta_t \\ \text{Observed Series} & = & \text{Permanent Trend} & + & \text{Cycle} \end{array} \tag{1}$$

However, as these constituent parts—trend and cycle—are not readily observed, any decomposition must necessarily be built on a conceptual artifact. Thus, any detrending method, must start out by somehow arbitrarily defining what shall be counted as *trend* and as *cycle*, before these elements can be estimated from the data.

The most common method used to extract the trend from a time series is the Hodrick-Prescott (HP) filter (Hodrick and Prescott, 1997). The HP filter extracts the trend, τ_t, by solving the following standard-penalty program:

$$\min_{\{\tau_t\}} \quad \underbrace{\sum_{t=1}^{T}(x_t - \tau_t)^2}_{\text{Goodness of Fit}} \; + \; \underbrace{\lambda \sum_{t=2}^{T-1}[(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2}_{\text{Penalty for Roughness}} \tag{2}$$

where the smoothing parameter λ controls the smoothness of the adjusted trend series, $\hat{\tau}_t$—*i.e.*, as $\lambda \to 0$, the trend approximates the actual series, x_t, while as $\lambda \to \infty$ the trend becomes linear.

While Hodrick and Prescott (1997) suggest values for λ, Marcet and Ravn (2003) recast the program (2) as a constrained minimization program to determine the value of λ endogenously. For annual data, λ should be between 6 and 7; see Ravn and Uhlig (2002), and Maravall (2004).

Note that the HP program (2) can be written more succintly as:[1]

$$\min_{\{\tau_t\}} \quad \sum_{t=1}^{T} \zeta_t^2 + \lambda \sum_{t=3}^{T} (\nabla^2 \tau_t)^2 \tag{3}$$

Which indicates that the HP filter attempts to maximize the fit of the trend to the series—*i.e.*, minimize the cycle component in (1)—while minimizing the changes in the trend's slope.

[1] Where $\nabla = (1 - B)$ is the standard differencing operator and B is the standard backshift (lag) operator, such that $B^j x_t = x_{t-j}$, and $\nabla x_t = x_t - x_{t-1}$. Also define the forward shifting operators: $F = B^{-1}$ and $\Delta = (1 - F)$.

1. The Generalized-Ridge Regression Form

Danthine and Girardin (1989) show that the solution to (2) is given by:

$$\hat{\tau} = [\mathbf{I} + \lambda \mathbf{K}'\mathbf{K}]^{-1}\mathbf{x} \tag{4}$$

where $\mathbf{x} = [x_1, \ldots, x_T]'$, $\tau = [\tau_1, \ldots, \tau_T]'$, $\mathbf{I}$ is a $T \times T$ identity matrix, and $\mathbf{K} = \{k_{ij}\}$ is a $(T-2) \times T$ matrix with elements given by

$$k_{ij} = \begin{cases} 1 & \text{if } i = j \text{ or } i = j+2, \\ -2 & \text{if } i = j+1, \\ 0 & \text{otherwise.} \end{cases}$$

that is:

$$\mathbf{K} = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{pmatrix}$$

Consider $\mathbf{x} = \mathbf{I}\tau + \zeta$, then a generalized-ridge regression rule would estimate the trend as

$$\hat{\tau} = [\mathbf{I}'\mathbf{I} + \lambda\mathbf{A}]^{-1}\mathbf{I}'\mathbf{x} = [\mathbf{I} + \lambda\mathbf{A}]^{-1}\mathbf{x}$$

where $\mathbf{A}$ is a symmetric, positive-definite matrix (Hoerl and Kennard, 1970). Making $\mathbf{A} = \mathbf{K}'\mathbf{K}$, it becomes apparent that the formulation (4) is a particular case of this approach. This implies that the HP trend can be interpreted as the fitted values of a ridge regression.

Relatedly, in a Bayesian framework, where $\zeta \sim N(0, \sigma^2)$, if we specifiy the natural conjugate prior $(\tau \mid \sigma) \sim N(0, (\sigma^2/\lambda)(\mathbf{K}'\mathbf{K})^{-1})$, then the posterior mean of τ is given by (4).

1.1. Example

As an illustration, let us consider, *e.g.*, $T = 5$, in which case:

$$\mathbf{I} + \lambda\mathbf{K}'\mathbf{K} = \begin{pmatrix} 1+\lambda & -2\lambda & -\lambda & 0 & 0 \\ -2\lambda & 1+5\lambda & 0 & -\lambda & 0 \\ -\lambda & 0 & 1+6\lambda & 0 & -\lambda \\ 0 & -\lambda & 0 & 1+5\lambda & 2\lambda \\ 0 & 0 & -\lambda & 2\lambda & 1+\lambda \end{pmatrix}$$

Making $\lambda = 7$, we compute $[\mathbf{I} + 7\,\mathbf{K}'\mathbf{K}]^{-1}$, and we obtain

$$\begin{pmatrix} \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \\ \hat{\tau}_4 \\ \hat{\tau}_5 \end{pmatrix} = \begin{pmatrix} \mathbf{0.644} & 0.375 & 0.156 & -0.014 & -0.161 \\ 0.375 & \mathbf{0.322} & 0.216 & 0.100 & -0.014 \\ 0.156 & 0.216 & \mathbf{0.254} & 0.216 & 0.156 \\ -0.014 & 0.100 & 0.216 & \mathbf{0.322} & 0.375 \\ -0.161 & -0.014 & 0.156 & 0.375 & \mathbf{0.644} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$$

A few things are worth noting:

(i) For each t, all the weights add to unity, and they do not depend on the data.

(ii) *Negative* weights occur for some observations (those more than 3 periods apart from t in this example).

(iii) The third observation—since it has equal number of observations before and after it—is the only one that will be filtered by a *symmetric* filter.

(iv) The endpoints will place a very large weight (0.649) on their observed values for determining the corresponding trend value, and the filter is one-sided at the endpoints.

(v) The trend for the observations next to the endpoints, however, will put a *larger* weight on the first and last observations (0.375) than on themselves (0.322).

The last two points, *(iv)–(v)*, illustrate the origin of the endpoint sample problem. When T gets revised, or when $T + k$ becomes available, there is a substantial effect on the estimates for τ_{T-1} and τ_T.

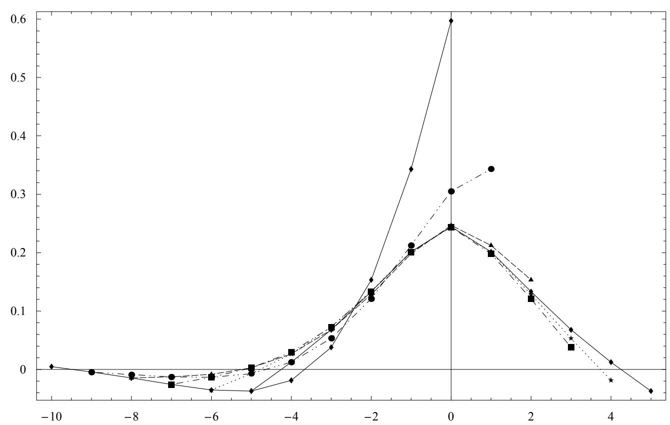

Fig. 1. HP Weights for T = 11, and λ = 7.
Horizontal axis: $k = 0, \pm 1, \pm 2, \ldots$; Vertical axis: weight on x_{t+k} for estimating τ_t

2. Model-Based Interpretations

In contrast to the original *ad hoc* formulation of the HP filter given by (2), the HP filter has different interpretations; it can be seen as a particular case of the Butterworth family of filters (Gómez, 1999), it can also be obtained in the context of an unobserved-components formulation (Harvey and Jaeger, 1993 and King and Rebelo, 1993), or as a Wiener-Kolmogorov filter (Kaiser and Maravall, 2001).

The unobserved-components (UC) representation is fairly general, as many popular decompositions, including the HP filter, can be formulated within its framework. The UC representation is generally given by

$$\text{Observed Series:} \quad x_t = \tau_t + \zeta_t \tag{5}$$

$$\text{Trend:} \quad \tau_t = \mu + \tau_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iidN(0, \sigma_\varepsilon^2) \tag{6}$$

$$\text{Cycle:} \quad \zeta_t \sim \text{ stationary and ergodic} \tag{7}$$

Often, the cycle, ζ_t, is assumed to have an ARMA(p, q) representation, $\varphi_p(B)\zeta_t = \vartheta_q(B)a_t$; the standard UC-ARMA decomposition assumes that the shocks to the trend and the cycle are uncorrelated—*i.e.*, it sets $\sigma_{a\varepsilon} = 0$. When this restriction is relaxed this decomposition leads to the Beverige-Nelson decomposition (Morley, Nelson and Zivot, 2003).

As noted by Harvey and Jaeger (1993) and King and Rebelo (1993), the HP filter can be interpreted as the optimal estimator in the UC model given by (5), with $\zeta_t \sim iidN(0, \sigma_c^2)$, and

$$\nabla^2\tau_t = \varepsilon_t, \qquad \varepsilon_t \sim iidN(0, \sigma_\varepsilon^2) \tag{8}$$

This formulation implies an IMA(2,2) model for x_t (Kaiser and Maravall, 1999), and therefore:

$$\nabla^2 x_t = \varepsilon_t + \nabla^2\zeta_t = (1 - \theta_1 B - \theta_2 B^2)b_t, \qquad b_t \sim iidN(0, \sigma_b^2) \tag{9}$$

As noted by Box and Jenkins (1976), optimal forecasts for an IMA(2,2) lie along a straight line, the level and the slope of which are continually updated as new data becomes available. (By contrast, in an IMA(1,1), new observations cannot update the slope, but only the level.)

The values of θ_1, θ_2 and σ_b^2 can be computed by factorizing the spectrum from the identity (9) (Maravall and Kaiser, 1999). Defining $\vartheta_{\text{HP}}(B) = (1 - \theta_1 B - \theta_2 B^2)$, the MMSE estimator for τ is given by the two-sided, symmetric and convergent linear filter:

$$\hat{\tau}_t = \frac{\sigma_c^2}{\sigma_b^2} \cdot \frac{\nabla^2\Delta^2}{\vartheta_{\text{HP}}(B)\vartheta_{\text{HP}}(F)} x_t$$

which requires the use of forecasts and backcasts to extend the series at the extremes of the observation interval.

3. References

Box, G.E.P. and G.M. Jenkins (1976), *Time Series Analysis, Forecasting and Control*, San Francisco CA: Holden-Day.

Gómez, Victor (1999), "Three Equivalent Methods for Filtering Nonstationary Time Series," *Journal of Business and Economics Statistics*, 17, 109–116.

Harvey, Andrew C. (1993), *Forecasting Structural Time Series and the Kalman Filter*, Cambdridge UK: Cambridge University Press.

Hodrick, R., and E.P. Prescott (1997), "Post-war Business Cycles: An Empirical Investigation," *Journal of Money, Credit, and Banking*, 29, 1–16

Kaiser, Regina and Agustín Maravall (2001), *Measuring Business Cycles in Economic Time Series*, Lecture Notes in Statistics 154, New York: Springer-Verlag.

King, Robert G. and Sergio Rebelo (1993), "Low Frequency Filtering and Real Business Cycles," *Journal of Economic Dynamics and Control*, 17, 207–233.

Marcet, Albert and Morten O. Ravn (2001), "The HP-Filter in Cross-Country Comparisons," Working Paper 588 UPF

Maravall, Agustín (1993), "Stochastic Linear Trends: Models and Estimators," *Journal of Econometrics*, 56, 5–37.

Maravall, Agustín (2004), Personal Communication.

Morley, James C., Charles R. Nelson and Eric Zivot (2003), "Why are the Beveridge-Nelson and Unobserved-Components Decomposition of GDP so Different?" *The Review of Economics and Statistics*, 85(2), 235–243.

Morten O. Ravn and Harald Uhlig (2002), "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations," *The Review of Economics and Statistics*, 84(2), pp. 371–375.

Prepared by Eduardo Ley. Run through pdfTEX on December 17, 2006 at 0:56 A.M.