Forecasting and Signal Extraction with Misspecified Models

Tommaso Proietti*
Dipartimento di Scienze Statistiche
Università di Udine

Abstract

The paper illustrates and compares estimation methods alternative to maximum likelihood, among which multistep estimation and leave-one-out cross-validation, for the purposes of signal extraction, and in particular the separation of the trend from the cycle in economic time series, and long-range forecasting, in the presence of a misspecified, but simply parameterised model. Our workhorse models are two popular unobserved components models, namely the local level and the local linear model. The paper introduces a metric for assessing the accuracy of the unobserved components estimates and concludes that cross-validation is not a suitable estimation criterion for the purpose considered, whereas multistep estimation can be valuable. Finally, we propose a local likelihood estimator in the frequency domain that provides a simple and alternative way of making operative the notion of emphasising the long-run properties of a time series.

Keywords: Business cycles, Unobserved components models, Cross-validation, Smoothing, Hodrick-Prescott filter, Multistep estimation.

*Address for Correspondence: Via Treppo 18, I-33100 Udine. e-mail: proietti@dss.uniud.it.

1 Introduction

This paper focusses on forecasting and signal extraction in the presence of model misspecification. We consider time series models which are both simple, in that their properties depend on a very limited set of parameter, and they generate predictors and signal extraction filters that are well understood. Even though they are suboptimal, it may be the case that they produce efficient forecasts and signal estimates, when the parameter estimation criteria are modified so as to enhance the features of interest for the specific problem at hand.

The kind of situation we have in mind arises in macroeconomic time series analysis when the trend is estimated by the popular Hodrick-Prescott (HP, 1999) filter or with a local linear trend model, in the presence of cyclical dynamics that are richer than those represented in the model, according to which the deviations from the trend are typically white noise. The signal extraction filter depends on a single smoothness parameter, whose maximum likelihood estimates (MLE) will result close to zero for most macroeconomic time series, as it is confirmed by empirical experimentation, implying a trend that absorbs most, if not all, of the variation in the data. For this reason, in empirical applications the smoothness parameter is calibrated, rather than estimated; frequency domain rules, linking it to a particular cut off frequency, see Gòmez (2001), can be viewed, from this perspective, as an attempt to extract meaningful cycles from a misspecified model.

Hence, in the presence of model misspecification, the estimates of unobserved components conditional on the maximum likelihood parameter estimates are not sensible. Residual diagnostics would point out the situation and a strategy would be to start the quest for an alternative model that improves the fit and accommodates those features that have been missed by the original specification, e.g. bringing in a cyclical component; however, this may be difficult to put into practice and may be costly, because more parameters have to be estimated.

This paper concentrates instead on the alternative strategy of keeping the model fixed and vary the estimation criterion, so as to elicit the features that are important for the purposes of signal extraction and long-range forecasting. That this strategy may be successful for the latter task is attested by the literature on multistep estimation, where the sum of squares of multistep forecast errors is the criterion function that is optimised; essential references are Cox (1961), Tiao and Xu (1993), Tiao and Tsay (1994), Clements and Hendry (1996), and Bhansali (2002). We aim at assessing whether there is a corresponding role that multistep estimation can play for signal extraction.

The aim of this paper is thus to discuss and compare estimation criteria alternative to MLE; focussing on two very popular unobserved components models, widely used for the decomposition of a time series, namely the local level and local linear trend model, we consider, along with multistep estimation (ME), cross-validation and local likelihood. Leave-one-out cross-validation (CV) is routinely used for parameter estimation in spline models (see, for instance, Green and Silverman, 1994) and in nonparametric regression; Kohn, Ansley and Wong (1991) and Kohn, Ansley and Tharm (1992) compared the performance of CV and MLE estimation of smoothing splines on the grounds of the capability of estimating a signal that is generated by given deterministic nonlinear function of time. Their concern is to determine the order of the spline and the estimation method that works best.

Our focus will be on the estimation of unobserved components, such as trends and cycles in macroeconomic time series; differently from the previous literature, we do not focus solely on multistep forecasting; moreover, we consider alternative criteria, such as cross-validation and we introduce local likelihood in the frequency domain.

In particular, we shall be concerned with the additive decomposition: $y_t = \mu_t + \epsilon_t$, where μ_t denotes the trend component and ϵ_t is the deviation from it, a stationary component. The notation $\tilde{\mu}_{t|t}$ will be used to denote the best linear estimator conditional on the true model and the observations up to and including time t; $\tilde{\mu}^*_{t|t}$ will denote the same component estimated from the misspecified model. Some analytic results will be valid based on the assumption of a doubly infinite sample; $\tilde{\mu}_{t|\infty}$ will denote the full sample estimate.

2 The local level model

The popular one-step-ahead predictor:

$$\tilde{y}_{t+1|t} = \lambda y_t + (1-\lambda)(y_t - \tilde{y}_{t|t-1}),$$

which yields an exponentially weighted moving average of the current and the past observations, is the optimal predictor for *local level* model:

$$\begin{array}{rclll} y_t & = & \mu_t + \epsilon_t, & t = 1, 2, \dots, T, & \epsilon_t \sim \mathrm{WN}(0, \sigma^2_\epsilon) \\ \mu_{t+1} & = & \mu_t + \eta_t, & & \eta_t \sim \mathrm{WN}(0, \sigma^2_\eta), \end{array} \quad (1)$$

where the disturbances are mutually uncorrelated; the reduced form is the IMA(1,1) model: $\Delta y_t = \xi_t + \theta \xi_{t-1}, \xi_t \sim \mathrm{WN}(0, \sigma^2)$. See Muth (1960), Cox (1961) and Harvey (1989).

Equating the autocovariance generating functions of Δy_t it is possible to establish that $\sigma^2_\eta = (1+\theta)^2\sigma^2$ and $\sigma^2_\epsilon = -\theta\sigma^2$. Hence, the structural model requires $\theta \leq 0$ and the signal to noise ratio, $q = \sigma^2_\eta/\sigma^2_\epsilon$, equals $-(1+\theta)^2/\theta$; moreover, $\lambda = 1 + \theta$.

In the steady state, the one-step-ahead prediction errors can be written as a linear combination of the original observations:

$$\tilde{\nu}_t = y_t - \tilde{y}_{t|t-1} = \left(1 - \frac{1+\theta}{1+\theta L}\right) y_t = \frac{1}{1+\theta L}\Delta y_t,$$

and they will be autocorrelated if Δy_t does not follow an MA(1) process with parameter θ. Only in the latter case they will be $\mathrm{WN}(0, \sigma^2)$.

The level predictions, filtered and smoothed estimates are, respectively:

$$\tilde{\mu}_{t+1|t} = \tilde{\mu}_{t|t} = \frac{(1+\theta)}{(1+\theta L)} y_t$$

$$\tilde{\mu}_{t|\infty} = \frac{(1+\theta)^2}{|1+\theta L|^2} y_t = \frac{(1+\theta)}{(1+\theta L^{-1})} \tilde{\mu}_{t|t};$$

these expressions follow from applying the Wiener-Kolmogorov prediction and signal extraction formulae, see Whittle (1983).

In the subsequent sections we shall be concerned with estimating θ: the model is then used for forecasting and, say, for detrending the series. If the model is correctly specified, then MLE is the most efficient option. The picture changes radically, however, when the model is misspecified. The next sections illustrate different ways of choosing θ and their virtues.

2.1 Method of moments estimators

The previous section illustrated that the parameter θ is essential in determining the weights that are attached to the observations for signal extraction and prediction. Let us consider now the case when ϵ_t is a stationary process with autocovariance generating function (ACGF) $\gamma_\epsilon(L) = \sum_{j=-\infty}^{\infty} \gamma_\epsilon(j) L^j$ and $\sigma_\epsilon^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \gamma_\epsilon(j) e^{-\imath\omega j} d\omega$, rather than white noise. A semiparametric estimator is readily available from the criterion of matching the long run properties of the misspecified ARIMA(1,1,0) model with those of the series under investigation.

Writing $\gamma(j) = \mathsf{E}(\Delta y_t \Delta y_{t-j})$, from the basic relationship

$$\sigma_\eta^2 = \gamma(0) + 2\sum_{j=1}^{\infty} \gamma(j) = g(0),$$

where $g(0)$ denotes the long run variance of Δy_t, and recalling that for a correctly specified model, $(1+\theta)^2\sigma^2 = \sigma_\eta^2$, it follows

$$\theta = \left(\frac{g(0)}{\sigma^2}\right)^{1/2} - 1 \tag{2}$$

The prediction error variance in the denominator can be expressed as a geometric average of the spectral generating function of Δy_t, that we denote with $g(\omega), \omega \in [0, \pi]$, using the Kolmogorov formula:

$$\sigma^2 = \exp\left[\frac{1}{\pi}\int_0^{\pi} \ln g(\omega) d\omega\right].$$

An estimate of θ can therefore be constructed from sample estimates of the prediction error variance and the long run variance; the latter is obtained via the kernel estimate:

$$\hat{g}(0) = \hat{\gamma}(0) + 2\sum_{j=1}^{l} w_l(j)\hat{\gamma}(j)$$

where l is the truncation lag and $w_l(j)$ is a suitable lag window, e.g. the Bartlett window, $w_l(j) = \frac{l-j+1}{l+1}$, in which case $\hat{\sigma}^2_\eta = \mathsf{Var}(\Delta_l y_t)/l$. The prediction error variance can be estimated by

$$\hat{\sigma}^2 = \exp\left[\frac{1}{T^*}\sum_{j=0}^{T^*-1} (\log 2\pi I(\omega_j)) + 0.57721\right]$$

(see Hannan and Nicholls, 1977), where $I(\omega_j)$ is the periodogram ordinate of $\Delta y_t, t = 1, \ldots, T^*$, evaluated at the Fourier frequency $\omega_j = \frac{2\pi j}{T^*}$, $j = 0, 1, \ldots, (T^* - 1)$.

Equation (2) expresses the fact that the estimator is based on the ratio between the estimate of the spectral density at at a particular frequency, namely the zero frequency, and the geometric average. The semiparametric estimator requires that long run predictability is greater than one-step-ahead predictability: $g(0) \leq \sigma^2$. A sufficient condition for the estimator to be feasible is that $g(\omega)$ is a minimum at the zero frequency, which guarantees $g(0) < \sigma^2$. This estimator is rarely feasible, as the long run variance should not be greater than the prediction error variance. This condition is stronger than that for the decomposability of the original model, which amounts to $g(0) < \gamma(0)$.

An alternative estimator uses the prediction error variance resulting from fitting the LLM, which amounts to $\sigma^2* = \gamma(0)/(1 + \theta^2)$; this is surely greater than σ^2, because we are fitting the wrong model, and makes it more likely that the estimated parameter is within the admissible range. Replacing into (2) and rearranging gives:

$$(1+\theta)^2 \left[\sum_{j=1}^{\infty} \gamma(j)\right] = \theta g(0). \tag{3}$$

This simple relation is at the basis of a semiparametric estimate of θ, which solves the quadratic equation $(1+\hat{\theta})^2 [\hat{g}(0) - \hat{\gamma}(0)] = 2\hat{\theta}\hat{g}(0)$, that uses sample estimates of the variance and the long run variance. Notice that we require the long run variance to be no greater than $\gamma(0)$ - this property is sometimes referred to as *mean reversion*.

The nonparametric estimators considered in this section use information about the zero frequency and compare it to a geometric or arithmetic average spectral average. In the next section we consider an alternative criterion, multistep estimation, that uses also information about $g(\omega)$ *around* the zero frequency. Loosely speaking, weaker forms of mean reversion are required.

2.2 Multistep Estimation

Multistep, or *adaptive, estimation* (ME) of the LLM has been considered by Cox (1961) Tiao and Xu (1993), Haywood and Tunnicliffe Wilson (1997), among others; see Bhansali (2002) for a comprehensive review of the approach. In particular, the relative efficiency of the multistep forecasts originating from the misspecified model, whose parameters are estimated minimising the variance of the l-step-ahead prediction errors, is not far from unity, for $l > 1$. Therefore, there exists a well established body of literature showing the merits of ME for the purpose of forecasting. Here we extend these results showing the properties of adaptive estimation for the purpose of signal extraction.

The l-step ahead forecast error arising form the IMA(1,1) model with MA parameter θ, here denoted by $\tilde{\nu}_{t+l|t}$, can be written as a linear combination of the current and past one-step-ahead forecast errors, $\tilde{\nu}_t$:

$$\tilde{\nu}_{t+l|t} = \left[1 + (1+\theta)L + (1+\theta)L^2 + \cdots + (1+\theta)L^{l-1}\right] \tilde{\nu}_t, \quad l > 1.$$

In terms of the observations, replacing $\tilde{\nu}_t = \Delta y_t/(1+\theta L)$, and rearranging,

$$\tilde{\nu}_{t+l|t} = v(L)\Delta y_t, \quad v(L) = S_{l-1}(L) + \frac{1}{1+\theta L} L^{l-1} \tag{4}$$

where $S_j(L)$ is the summation operator involving j consecutive terms, $S_j(L) = 1+L+L^2+\cdots+L^{j-1}$. An alternative expression is $v(L) = [(1+\theta)S_l(L)-\theta]/(1+\theta L)$.

Multistep estimation determines θ as the minimiser of

$$ME(\theta, l) = \mathsf{Var}(\tilde{\nu}_{t+l|t}) = \frac{1}{\pi}\int_0^\pi |v(e^{-\imath\omega})|^2 g(\omega) d\omega, \tag{5}$$

where $|v(e^{-\imath\omega})|^2 = v(e^{-\imath\omega})v(e^{\imath\omega})$ is the squared gain of the filter $v(L)$, and $g(\omega)$ is the spectral generating function of Δy_t.

ME can thus be viewed as minimising the variance of a *filtered* series. The results will depend on the properties of the series, an important feature being its order of integration. The first panel of figure 1 plots the squared gain of the filter $v(L)\Delta$ for $\theta = -0.8$ and $l = 1, 2, 5, 10$; the gain is zero at the zero frequency and this implies that if y_t is stationary the zero frequency is not informative on θ. On the other hand, for difference stationary series, the plot of the transfer function of $v(L)$ in the second panel shows that the filter emphasises the spectral density around the zero frequency, see Haywood and Tunnicliffe Wilson (1997); furthermore, the concentration of power around the zero frequency increases with l. This suggests that for $I(1)$ series the ME estimate with l large will give more relevance to the long run features of the series.

[Figure 1 about here]

The rest of this section aims at showing the connection of the ME estimator of θ

with the long run properties of the series. Using (11),

$$\begin{aligned}\mathsf{Var}(\tilde{\nu}_{t+l|t}) &= V(l) + \tfrac{1}{1-\theta^2}\left\{\gamma(0) + 2\sum_{j=1}^{\infty}(-\theta)^j\gamma(j) + 2(1-\theta)\sum_{j=1}^{l-2}\left[1-(-\theta)^j\right]\gamma(j) + \right.\\ &\left. 2(1-\theta)\left[1-(-\theta)^{l-1}\right]\sum_{j=l-1}^{\infty}(-\theta)^{j-l+1}\gamma(j)\right\},\end{aligned}$$

where

$$V(l) = (l-1)\gamma(0) + 2\sum_{j=1}^{l-2}(l-j-1)\gamma(j)$$

is the leading term and is invariant to θ.

As the forecast horizon gets bigger the formula tends to

$$\lim_{l\to\infty}\left[\mathsf{Var}(\tilde{\nu}_{t+l|t}) - V(l)\right] = \frac{1}{1-\theta^2}\left\{g(0) - 2\theta\sum_{j=1}^{\infty}\left[1-(-\theta)^j\right]\gamma(j)\right\}.$$

Differentiating with respect to θ and setting the derivative equal to zero, which is the first order condition for a minimum, yields the nonlinear equation:

$$(1+\theta)^2 A(\theta) = \theta g(0) + B(\theta), \tag{6}$$

$$\begin{aligned} A(\theta) &= \sum_{j=1}^{\infty}\gamma(j) + \sum_{j=1}^{\infty}(-\theta)^j(j-1)\gamma(j), \\ B(\theta) &= 2\theta\left[A(\theta) - \sum(-\theta)^{j-1}j\gamma(j)\right].\end{aligned} \tag{7}$$

If y_t is stationary, with autocovariances $\mathsf{Cov}(y_t, y_{t-j}) = \gamma^*(j)$, then $g(0) = 0$ and the unique solution is $\theta = -1$: the EWMA predictor converges to the time average as the forecast horizon increases. This is so since $B(-1) = 0$, as it can be easily checked, and $A(-1) = \sum_j j\gamma(j) = \gamma^*(1) - \gamma^*(0)$, which is different from zero; the last result uses the well known identity $\gamma(j) = 2\gamma^*(j) - \gamma^*(j-1) - \gamma^*(j+1)$.

If Δy_t is stationary, θ will converge to a finite value, greater than -1, that reflects the persistence of the process, although it will not exactly satisfy (3). This is so because the multistep filter acts as a lowpass filter exploiting the information around the zero frequency as well. Increasing l enhances the low-pass nature of the filter, that however will render the criterion function flatter with respect to θ; the multistep filter becomes so concentrated that a more limited frequency band is considered with the consequence that we use less information for the estimation of θ_2. The main point is that a negative θ is available under less stringent conditions than those embedded in (3).

2.3 Cross-validation

The smoothed estimates of the irregular component in the LLM are $\tilde{\epsilon}_{t|\infty} = \sigma^2_\epsilon u_t$, where u_t is known as a smoothing error (de Jong, 1988, Kohn and Ansley, 1989, Koopman, 1993). In the steady state, the latter is provided by:

$$u_t = \frac{|1-L|^2}{\sigma^2|1+\theta L|^2} y_t,$$

where $|1-L|^2 = (1-L)(1-L^{-1})$, $|1+\theta L|^2 = (1+\theta L)(1+\theta L^{-1})$, and it should be noticed that the impulse response function of the filter applied to the observations is provided by the ACGF of the inverse ARMA(1,1) model $(1+\theta L)y^*_t = \Delta\xi^*_t$. If $|\theta| < 1$ the variance of u_t, denoted $M = \mathsf{Var}(y^*_t)$, is given by the expression $M = 2\sigma^2/(1-\theta)$ and as shown by de Jong (1988), the interpolation error is a simple function of u_t scaled by the inverse of its variance:

$$y_t - \mathsf{E}[y_t|Y_{\setminus t}] = M^{-1}u_t.$$

Note that for $\theta = 0$ we get the RW interpolation formula $\mathsf{E}[y_t|Y_{\setminus t}] = 0.5(L+L^{-1})y_t$.

Cross-validation (CV) is based on the minimisation of the sum of squares of

$$y_t - \mathsf{E}[y_t|Y_{\setminus t}] = \frac{u_t}{M} = \frac{1}{2}\frac{1-\theta}{|1+\theta L|^2}|1-L|^2 y_t = u(L)\Delta y_t$$

where

$$u(L) = \frac{1}{2}\frac{1-\theta}{|1+\theta L|^2}(1-L^{-1}).$$

The cross-validatory estimate of the parameter θ is thus the minimiser of the function:

$$CV(\theta) = \frac{1}{\pi}\int_0^\pi |u(e^{-\imath\omega})|^2 g(\omega)d\omega \tag{8}$$

where, as before, $g(\omega)$ is the spectral generating function of Δy_t and $|u(e^{-\imath\omega})|^2$ is the squared gain of $u(L)$.

The bottom panels of fig. 1 depict the squared gains of the filters $u(L)\Delta$ and $u(L)$. A relevant difference with respect to ME arises for first order integrated y_t, for which $CV(\theta)$ does not use the zero frequency.

3 Evaluating the performance of the approximating model

Since the approximating model is used for signal extraction and forecasting, there are two aspects that need to be evaluated. As far as the second is concerned, the relative forecast accuracy of the approximating model can be assessed by comparing its l-step-ahead forecast error variance, evaluated at the minimiser of (5), $ME(\hat{\theta}, l)$, with that of the optimal forecast, under the true model, denoted $\mathsf{Var}(\nu_{t+l|t})$:

$$RelEff(l) = \frac{ME(\hat{\theta}, l)}{\mathsf{Var}(\nu_{t+l|t})}$$

The assessment of the performance concerning the estimation of unobserved components relies on the availability of a doubly infinite sample. Suppose the true generating model is $y_t = \mu_t + \epsilon_t$ with orthogonal components, so that, if $g_\mu(L)$ and $g_\epsilon(L)$ denote the autocovariance generating functions of the trend and the cycle, $g_y(L) = g_\mu(L) + g_\epsilon(L)$.

Let also $w_{\mu*}(L)$ and $w_{\epsilon*}(L)$ denote the Wiener-Kolmogorov signal extraction filters for the approximating model $y_t = \mu_t^* + \epsilon_t^*$, where

$$w_{\mu*}(L) = \frac{g_{\mu*}(L)}{g_{\mu*}(L) + g_{\epsilon*}(L)}, \qquad w_{\epsilon*}(L) = \frac{g_{\epsilon*}(L)}{g_{\mu*}(L) + g_{\epsilon*}(L)}.$$

THEOREM: The ACGF of the unobserved components estimation error, $g_e(L)$, can be expressed in terms of the squared gains of the signal extraction filters and the ACGF of the true components in the following manner:

$$g_e(L) = g_\mu(L)|w_{\epsilon*}(L)|^2 + g_\epsilon(L)|w_{\mu*}(L)|^2. \tag{9}$$

The minimum is

$$\frac{g_\mu(L)g_\epsilon(L)}{g_y(L)}.$$

Proof: the ACGF of $e_t = \epsilon_t - \tilde{\epsilon}_t^*$ is

$$\begin{aligned} g_e(L) &= g_\epsilon(L) + |w_{\epsilon*}(L)|^2 g_y(L) - g_\epsilon(L)[w_{\epsilon*}(L) + w_{\epsilon*}(L^{-1})] \\ &= g_\epsilon(L)\,[1 - w_{\epsilon*}(L)]\,[1 - w_{\epsilon*}(L^{-1})] + g_\mu(L)|w_{\epsilon*}(L)|^2 \\ &= g_\epsilon(L)|w_{\mu*}(L)|^2 + g_\mu(L)|w_{\epsilon*}(L)|^2 \end{aligned}$$

as $\tilde{\epsilon}_t^* = w_{\epsilon*}(L)y_t$ $\tilde{\mu}_t^* = w_{\mu*}(L)y_t = [1 - w_{\epsilon*}(L)]\,y_t$.

When the model is correctly specified,

$$w_{\mu*}(L) = \frac{g_\mu(L)}{g_y(L)}, \qquad w_{\epsilon*}(L) = \frac{g_\epsilon(L)}{g_y(L)}$$

and, therefore, the minimum stated above is achieved. The latter is the estimation error ACGF as given in Whittle (1983, p. 58). Notice that (9) is symmetric, that is $e_t = -(\mu_t - \tilde{\mu}_t^*)$■

The estimation error MSE is thus

$$\mathsf{MSE}(e_t) = \frac{1}{\pi}\int_0^\pi g_e(\omega)d\omega;$$

this can be evaluated via numerical integration when the true generating model and filters are known. For the LLM the estimation error ACGF is:

$$g_\epsilon(L) = \frac{-\theta(1+\theta)^2}{|1+\theta L|^2}\sigma^2.$$

4 Illustrative examples

This section illustrates the behaviour of ME and CV estimates of the parameter θ of the LLM in the traditional case when the true data generating process is a stationary autoregression, and in a less explored, but interesting case, when the true model is ARIMA(1,1,0).

Multistep estimation of the LLM when the true model is a stationary AR(1) process, $y_t = \phi y_{t-1} + \xi_t$, $\xi_t \sim \mathrm{WN}(0, \sigma^2)$, $|\phi| < 1$, has been investigated by Cox (1961). Figure 2 shows the value of θ minimising[1] the variance of the l-step-ahead forecast for $l = 1, 2, 5, 10$ and the CV estimates for different values of ϕ in the range (-1,1). The plot reveals that $\hat{\theta}$ tends to -1 as l increases: for l odd, a negative θ can only arise for $\phi > 0$; on the other hand, if l is even, $-1 < \theta < 0$ can also arise for negative values of ϕ (this is discordant from Cox (1961) statement that the optimal ME estimate is $\theta = -1$ for $\phi < 1/3$). As the AR parameter tends to 1, the optimal θ is zero (RW predictions). Cross-validation estimates are close to ME with $l = 1$.

[Figure 2 about here]

Figure 3 shows the relative forecast efficiency of ME estimates, given by the ratio of $ME(\hat{\theta}, l)$ to the mean square forecast error of the true model ($\sigma^2 = 1$), $(1-\phi^{2l})/(1-\phi^2)$. It should be noticed that unless ϕ is close to 1 multistep LLM forecasts for l even are almost as efficient as the true model's forecasts.

[Figure 3 about here]

Consider now the ARIMA(1,1,0) process

$$\Delta y_t = \phi \Delta y_{t-1} + \xi_t, \xi_t \sim \mathrm{WN}(0, \sigma^2)$$

As shown in figure 4, the ME and CV estimates of θ are zero when ϕ is positive. The plot confirms that CV estimates are closer to ME with $l = 1$. For l even and a negative ϕ, ME is almost as efficient as the correct model forecasts even at short horizons; the performance deteriorates for positive ϕ.

[1]The values are obtained by evaluating numerically the integrals $MS(\theta, l)$ in (5) and $CV(\theta)$ in (8). The computations were carried out in Ox3.3, see Doornik (2001).

[Figure 4 about here]

When ϕ is negative, Proietti and Harvey (2000) showed that the process can be decomposed into a RW trend plus a stationary AR(1) component, $y_t = \mu_t + \epsilon_t$, where

$$\Delta\mu_t = \eta_t,\ \eta_t \sim \text{WN}\left(0, \frac{\sigma^2}{(1-\phi)^2}\right), \quad \epsilon_t = \phi\epsilon_t + \kappa_t,\ \ \kappa_t \sim \text{WN}\left(0, -\frac{\phi\sigma^2}{(1-\phi)^2}\right)$$

The Wiener-Kolmogorov estimator of the trend is $\tilde{\mu}_{t|\infty} = (1-\phi)^{-2}|1-\phi L|^2 y_t$, see Proietti and Harvey (2000), and thus it is provided by a filter with finite impulse response. The trend extraction filter for the LLM has an infinite impulse response and the the issue is whether efficiency in multistep forecast provides a clear guidance also over that concerning the estimation of unobserved components.

The answer is provided by the right panel of figure 5, which displays, on a logarithmic scale, the measure (9) divided by its minimum, that is attained when the model is correctly specified. While it is confirmed that the efficiency increases as ϕ decreases, the efficiency for l odd is greater than for l even.

This example shows that there is a difference between the performance in forecasting and in the estimation of unobserved components. Moreover, cross-validation tends to be less efficient than multistep estimation with respect to the estimation of unobserved components.

[Figure 5 about here]

5 The local linear trend model

Another popular model for forecasting and signal extraction is the local linear trend model (LLTM), see Harvey (1989), West and Harrison (1997) and Young and Pedregal (1999), which is formulated as follows:

$$\begin{array}{rcllllll} y_t & = & \mu_t & + & \epsilon_t, & & \epsilon_t \sim \text{WN}(0, \sigma^2_\epsilon), & t = 1, 2, \dots, T, \\ \mu_{t+1} & = & \mu_t & + & \beta_t & + \ \eta_t, & \eta_t \sim \text{WN}(0, \sigma^2_\eta), & \\ \beta_{t+1} & = & & & \beta_t & + \ \zeta_t, & \zeta_t \sim \text{WN}(0, \sigma^2_\zeta), & \end{array} \quad (10)$$

The model for the trend features a stochastic drift.

The reduced form is the IMA(2,2) model:

$$\Delta^2 y_t = (1 + \theta_1 L + \theta_2 L^2)\xi_t, \quad \xi_t \sim \mathrm{WN}(0, \sigma^2)$$

Equating the ACGF of the structural form with that of the reduced form yields:

$$\sigma^2_\eta = -[\theta_1(1 + \theta_2) + 4\theta_2]\sigma^2, \quad \sigma^2_\zeta = \theta(1)^2\sigma^2, \quad \sigma^2_\epsilon = \theta_2\sigma^2.$$

These relations determine the region of admissible MA parameter space; in particular, $\sigma^2_\epsilon \geq 0$ requires $\theta_2 \geq 0$; furthermore, $\sigma^2_\eta \geq 0$ if and only if

$$\theta_1 \leq -\frac{4\theta_2}{1 + \theta_2}.$$

When the equality $\theta_1 = -4\theta_2/(1 + \theta_2)$ holds, $\sigma^2_\eta = 0$, $\theta(1) = \frac{(1-\theta_2)^2}{1+\theta_2}$, and the signal to noise ratio $\sigma^2_\zeta/\sigma^2_\epsilon$ is a function of θ_2 alone, being equal to $(1 - \theta_2)^4/[(1 + \theta_2)^2\theta_2]$.

The forecast function is $\tilde{y}_{t+l|t} = \tilde{\mu}_{t|t} + l\tilde{\beta}_{t|t}$, where the steady state recursions for $\tilde{\mu}_{t|t}$ and $\tilde{\beta}_{t|t}$ are equivalent to those of the Holt-Winters' forecasting technique:

$$\begin{array}{rcccccc} \tilde{\mu}_{t|t} & = & \tilde{\mu}_{t-1|t-1} & + & \tilde{\beta}_{t-1|t-1} & + & \lambda_0\nu_t \\ \tilde{\beta}_{t|t} & = & & + & \tilde{\beta}_{t-1|t-1} & + & \lambda_0\lambda_1\nu_t \end{array}$$

with

$$\lambda_0 = 1 - \theta_2, \quad \lambda_0\lambda_1 = \theta(1).$$

The smoothing constants λ_0 and λ_1 are both in the range (0,1), as $\sigma^2_\eta \geq 0$ implies $0 < \theta(1) < 1 - \theta_2$.

As shown in Proietti (2002), the steady state weights attributed to the observations for deriving the filtered and smoothed estimates can be derived from the expressions:

$$\tilde{\mu}_{t|t} = \left[\frac{1 - \theta_2}{\theta(L)}\Delta + \frac{\theta(1)}{\theta(L)}L\right] y_t, \quad \tilde{\beta}_{t|t} = \frac{\theta(1)}{\theta(L)}\Delta y_t.$$

Note that the weights are less than 1 in modulus and sum up to 1 and to 0 respectively for the level and the slope.

The LLTM nests several special cases of interest:

- When $\sigma^2_\epsilon = 0$, $\theta_2 = 0$ and the reduced form is IMA(2,1). The irregular is absent and the trend is coincident with the observations. If further $\sigma^2_\eta = 0$, the series is an integrated random walk, $\Delta^2 y_t = \xi_t$.

- When $\sigma^2_\zeta = 0$ the slope is constant, $\beta_t = \beta$. This in turn implies $\theta(L) = \Delta(1+\theta L)$, $\theta(1) = 0$, and an IMA(1,1) reduced form; the formulae for signal extraction were as given for the LLM (section 2), whereas the forecast function is linear in the forecast horizon, $\tilde{y}_{t+l|t} = \tilde{\mu}_{t|t} + l\beta$, with constant slope.

- When $\sigma^2_\eta = 0$, the model generates the celebrated Hodrick-Prescott filter, with smoothing parameter $\lambda = \sigma^2_\epsilon/\sigma^2_\zeta = \frac{\theta_2(1+\theta_2)^2}{(1-\theta_2)^4}$.

In the last case, which we label HP henceforth, the filtered and the smoothed estimates of the trend and the irregular are respectively:

$$\tilde{\mu}_{t|t} = \frac{1-\theta_2}{\theta(L)}\left[\Delta + \frac{1-\theta_2}{1+\theta_2}L\right]y_t, \quad \tilde{\mu}_{t|\infty} = \frac{\theta(1)^2}{|\theta(L)|^2}y_t$$

$$\tilde{\epsilon}_{t|t} = \frac{\theta_2}{\theta(L)}\Delta^2 y_t = \theta_2\tilde{\nu}_t, \quad \tilde{\epsilon}_{t|\infty} = \frac{\theta_2}{|\theta(L)|^2}|1-L|^4 y_t = \frac{(1-L^{-1})^2}{\theta(L^{-1})}\tilde{\epsilon}_{t|t}$$

where $\tilde{\nu}_t = \theta(L)^{-1}\Delta^2 y_t$ are the innovations.

These expressions are easily derived respectively from the steady state state recursions and from straight application of the Wiener-Kolmogorov filter, see Whittle (1983). Moreover, it should be noticed that under the HP restriction we have the nice decomposition: $|\theta(L)|^2 = \theta(1)^2 + \theta_2|1-L|^4$.

5.1 Multistep estimation of the LLTM

The l-step-ahead prediction error for the LLM can be written:

$$\tilde{\nu}_{t+l|t} = \left[1 + \vartheta_1 L + \vartheta_2 L^2 + \cdots + \vartheta_{l-1}L^{l-1}\right]\tilde{\nu}_t, \quad \vartheta_j = (1-\theta_2) + j\theta(1)$$

where in the steady state $\theta(L)\tilde{\nu}_t = \Delta^2 y_t$; thus, in terms of the observations:

$$\tilde{\nu}_{t+l|t} = \upsilon(L)\Delta^2 y_t, \quad \upsilon(L) = \frac{\vartheta(L)}{\theta(L)}. \tag{11}$$

In the HP case a single parameter, θ_2, determines the properties of the multistep filter. The closer is θ_2 to zero, the smaller the variation attributed to the irregular component; on the contrary, as θ_2 approaches 1, the trend will be more stable and more variation will be absorbed by the irregular component. The plots in the top and central rows of figure 6 present the squared gain of the filters $v(L)\Delta^2$, $v(L)\Delta$ and $v(L)$ for θ_2 equal to 1/4 and 1/2 and $l = 1$ and $l = 5$. The filters apply respectively when y_t, Δy_t and $\Delta^2 y_t$ are stationary and give some clue over the nature of inferences made by multistep estimation. When the series is integrated of the second order the filter gives more weight to the long run frequency, whereas in the previous two cases the gain is zero at the zero frequency. The multistep filter becomes more selective as θ_2 increases.

[Figure 6 about here]

5.2 Cross-validation for the LLTM

For the HP model, the smoothed estimates of the irregular component are $\tilde{\epsilon}_{t|\infty} = \sigma^2_\epsilon u_t$ where

$$u_t = \frac{|1-L|^4}{\sigma^2|\theta(L)|^2} y_t$$

is the smoothing error. Again the ACGF of u_t coincides with the inverse ACGF of the LLTM. Moreover, it can be shown that that the variance of u_t is, in the steady state[2]:

$$M = \sigma^{-2}\left[1 + n\frac{(1-\theta_2)^2}{1+\theta_2}(5+\theta_2)/2\right], \quad n = \left[1 + \theta_2\frac{1-3\theta_2}{1+\theta_2} - \frac{1}{2}(1-\theta_2)^2\right]^{-1}$$

CV estimation of θ_2 is based upon the minimisation of the variance of u_t/M; defining

$$u_0(L) = \frac{|1-L|^4}{M\sigma^2|\theta(L)|^2}, \;\; u_1(L) = \frac{|1-L|^2}{M\sigma^2|\theta(L)|^2}(1-L^{-1}), \;\; u_2(L) = \frac{(1-L^{-1})^2}{M\sigma^2|\theta(L)|^2},$$

[2]The proof is available from the author.

we have that the interpolation error is obtained applying the filter $u_i(L), i = 0, 1, 2$ respectively to y_t, Δy_t, and $\Delta^2 y_t$. Hence, $u_i(L)$ provides the relevant transformation for series integrated of the i-th order. For instance, if $g(\omega)$ denotes the spectral generating function of $\Delta^2 y_t$, then the cross-validatory estimator of θ_2 minimises

$$\frac{1}{\pi}\int_0^{\pi} |u_2(\omega)|^2 g(\omega) d\omega$$

The bottom row of figure 6 displays the squared gain $|u_i(\omega)|^2, i = 0, 1, 2,$ for $\theta_2 = 0.25, 0.5$; the plots reveal an interesting feature: while the gain of $u_0(L)$ and $u_1(L)$ is similar to the corresponding one-step-ahead filter (compare the top and bottom plots), there is a significant difference between $|u_2(\omega)|^2$ and $|v(L)|^2$ for $l = 1$, both of which apply to $I(2)$ series. As a matter of fact, the former is not a low-pass filter, strictly speaking, as it annihilates the zero frequency; rather, it has the nature of a cyclical band-pass filter with a spectral peak depending on the parameter θ_2.

The plot also illustrates that if the order of integration of the series coincides with that of the approximating model, the LLTM with $\sigma^2_\eta = 0$ in our case, the multistep filter has a low-pass nature; otherwise, it emphasises the cyclical or the high frequencies.

6 Illustrative examples

The first illustration deals with an application of the LLTM with constant drift ($\sigma^2_\zeta = 0$) for the extraction of the trend component from the logarithms of Italian GDP, plotted in figure 7. Since $\theta(L) = \Delta(1 + \theta L)$, the approximating model has reduced form IMA(1,1), $\Delta y_t = \beta + (1 + \theta L)\xi_t, \xi_t \sim \mathrm{WN}(0, \sigma^2)$. If Gaussian disturbances are assumed for this series, maximum likelihood estimation (MLE) yields an estimate of σ^2_ϵ that is practically zero, so that all the variation is absorbed by the trend component; see the top row of figure 7. Hence, the decomposition is based on a random walk model with drift, but this is clearly inadequate for the series, as the statistical significance and the pattern of residual autocorrelation show. The latter has a pseudo-cyclical behaviour that is suggestive of the presence of a cyclical component. Minimising the variance of the l-step-ahead prediction errors yields different results: as the plots at the bottom illustrate - they refer to the signal to noise ratio and the implied θ estimates as a function of the forecast horizon - the variance of the irregular component grows with l, relative to that of the changes in the trend. The central panels display the smoothed estimates of the components[3] for $l = 5$. The estimated irregular has now larger amplitude and richer dynamics than white noise. Cross-validation yields the same estimate as maximum likelihood.

[Figure 7 about here]

As a second illustration we consider the generating process $y_t = \mu_t + \epsilon_t$, where μ_t is alternatively a random walk (RW), as in (1), or an integrated random (IRW), as in (10) with the HP restriction, and ϵ_t is is the cyclic process (see Harvey, 1989):

$$\begin{bmatrix} \epsilon_{t+1} \\ \epsilon^*_{t+1} \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon^*_t \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa^*_t \end{bmatrix}, \tag{12}$$

where $\kappa_t \sim \text{NID}(0, \sigma^2_\kappa)$ and $\kappa^*_t \sim \text{NID}(0, \sigma^2_\kappa)$ are mutually uncorrelated, and uncorrelated with the trend disturbances. Hence, $\epsilon_t \sim$ARMA(2,1) and $\sigma^2_\epsilon = \sigma^2_\kappa/(1-\rho^2)$.

We set $\rho = 0.9$ and $\lambda_c = \pi/8$, corresponding to a period of 4 years of quarterly observations and we consider values of the signal to noise (SN) ratio ranging from 10^{-4} to 10^4 (when the trend is a RW we refer to $\sigma^2_\eta/\sigma^2_\epsilon$; when it is an IRW we refer to $\sigma^2_\zeta/\sigma^2_\epsilon$).

We start with the case when μ_t is a RW (the true model generating y_t is the sum of a RW and a stationary ARMA(2,1) cycle) and the approximating model is the LLM; the source of misspecification is the representation adopted for the

[3]The MLE and the smoothed estimates of the components are obtained using SsfPack (beta) v. 3.0 by Koopman *et al.* (1999).

cyclical component by the approximating model, which is white noise, rather than ARMA(2,1). The multistep estimates of the parameter θ are displayed in the first graph of figure 8 as a function of $\log_{10} \sigma^2_\eta/\sigma^2_\epsilon$ and of the forecast horizon. For $l = 1$, $\hat{\theta} = 0$ (RW prediction), regardless of the SN ratio, but for $l > 1$ $\hat{\theta}$ is around -1 (mean prediction) when the cycle is the dominant source of variation, and gradually moves to 0 as the reverse is true. The periodicity in the estimates as l varies reflects the fact that the true model is cyclical, but it is amplified or reduced by the multistep filter. The natural logarithms of the relative forecast efficiency is plotted in the second graph of the first row. Understandably, this is close to zero (the approximating model is fully efficient) when the SN ratio is high or low.

As far as the estimation of the trend and the cycle is concerned, the performance is poor for short horizons when the SN ratio is low. It improves with l, although the relative efficiency ratio it is still large for values of SN in the range $[10^{-2}, 1]$. The cross-validatory estimates are $\hat{\theta} = 0$ and are coincident with the ME with $l = 1$; as such they have the same efficiency in the estimation of unobserved components.

When the true μ_t is an IRW (centre row of figure 8) the multistep estimates of θ_2, the core parameter of the HP approximating model, behave like those of θ for the previous case, on a reverse scale; a difference arise, however, for $l = 1$ as positive and small estimates arise for low SN ratios. The performance in forecasting is satisfactory for low values of the SN ratio, although the signal extraction efficiency is usually so. Cross-validation, on the other hand, produces estimates that are closer to zero, implying an IRW representation for the series; as a consequence, the signal extraction efficiency is very low when the cycle is the dominant source of variation. This fact is a consequence of the high-pass nature of the cross-validation filter.

[Figure 8 about here]

The bottom panel considers instead the case when the true model is made up of a RW trend and a stationary cycle, but the LLTM with the HP restriction is fitted.

Hence, there is a second source of misspecification, which concerns the order of integration of the trend. Note that in this case the relative unobserved components estimation error diverges, as can be seen from (9), which shows that the spectral generating function is unbounded at the zero frequency; nevertheless, the relative forecast MSE is finite, and it is displayed in the last figure. The multistep estimates of θ_2 converge to 1 as l increases, implying a deterministic linear predictor ($\sigma^2_\zeta = 0$). As a result, the forecast efficiency is poor also when the trend is the dominant source of variation.

The reason why we restrict our attention to the HP case, rather than to the unrestricted LLTM, lies in the fact that the HP filter is much used and abused for the estimation of the trend in economic time series; see Pedregal and Young (2001) and the references therein for a thorough account of this point.

7 Local likelihood

As we saw in section 2.2, multistep estimation emphasises the long-run features of the series in the estimation of the parameters of a given model. We now propose and evaluate an alternative estimation method, which we call local likelihood, that has the same objective of giving more weight to particular aspects of the series. The natural set up for our purposes is the frequency domain.

Suppose that the series is difference stationary and that the approximating model is the LLM of section 2; given the availability of T^* observations $\Delta y_t, t = 1, 2, \dots, T^*$, let us denote the Fourier frequencies by $\omega_j = \frac{2\pi j}{T^*}$, $j = 0, 1, \dots, (T^* - 1)$. Apart from a constant, the Whittle's likelihood is then defined as follows (Harvey, 1989):

$$\text{loglik} = -\frac{1}{2}\sum_{j=0}^{T^*-1}\left[\log g^*(\omega_j) + 2\pi\frac{I(\omega_j)}{g^*(\omega_j)}\right] \tag{13}$$

where $g^*(\omega_j) = g^*(e^{-\imath\omega_j})$ denotes the spectral generating function of the stationary representation of the approximating model evaluated at frequency ω_j, that is

$g^*(\omega_j) = \sigma^2(1 + \theta^2 + 2\theta \cos \omega_j)$, and $I(\omega_j)$ is the periodogram:

$$I(\omega_j) = \frac{1}{2\pi}\left[c_0 + 2\sum_{\tau=1}^{T^*-1} c_\tau \cos(\omega_j \tau)\right]$$

where c_τ denotes the sample autocovariance at lag τ of Δy_t.

A local likelihood estimate is intended to give more weight to the frequencies around a target frequency, ω_0; the objective function can be written as:

$$\text{loglik}(\omega_0) = -\frac{1}{2}\sum_{j=0}^{T^*-1} w_j \left[\log g^*(\omega_j) + 2\pi \frac{I(\omega_j)}{g^*(\omega_j)}\right],$$

where $w_j = K(\omega_j - \omega_0)$ is a weighting function depending on the distance from the target frequency. In multistep estimation the kernel is automatically provided by the forecast function of the approximating model and depends on its parameters. In the local likelihood approach the kernel can be made independent of the approximating model. For instance, if our interest lies in long range forecasting and in the estimation of long-run trends then ω_0 is the zero frequency ; we may thus reduce the weight attached to the fit of high frequency periodogram ordinates, the latter being influenced by uninteresting fluctuations, in this respect, such as trading day variation and other short lived components.

Another justification for downweighting the high frequencies arises when pre-filtering by moving averages or temporal aggregation have taken place, so that the original amplitude of the frequency components in the series has been modified to an extent that the conditions for an orthogonal trend-irregular decomposition are no longer met.

One option is to consider the weighting function:

$$w_j = \begin{cases} 1, & \omega_j \leq \omega_c \\ 0, & \omega_j > \omega_c \end{cases}$$

here ω_c is the cutoff frequency and the kernel is the uniform kernel in the interval $[0, \omega_c]$.

We illustrate this approach with respect to the problem of extracting the trend from the logarithms of the Italian and U.S. quarterly GDP, using the LLTM with constant drift as the approximating model. The maximiser of (13) is $\hat{\theta} = 0$ in both cases, implying that the trend is coincident with the observations; the usual diagnostics highlight the presence of misspecification. The periodogram of Δy_t, displayed in the right plots of figure 9 only for the frequency range $[0, \pi]$ due to its symmetry around π, is interpolated by a constant spectrum; the resulting trend extraction filter uses only the current observation with unit weight.

If we downweight the high frequencies we get local likelihood estimates that move away from zero, imply smoother trends. This fact is illustrated by figure 9: the plots on the left hand side display the estimated θ values for cutoff frequencies in the range $\pi/10$ (corresponding to a period of 5 years) and $2\pi/5$ (corresponding to a period of 5 quarters). Those on the right hand side display, along with the periodogram of Δy_t, the parametric spectral density implied by the LLTM with constant drift, that has been fitted using the periodogram up to the cutoff frequency $\omega_c = \pi/4$ (corresponding to a period of 2 years).

In the Italian GDP case, when $.5 < \omega_c < 1$, $\hat{\theta}$ moves away from zero; the estimates are highly influenced by a single periodogram ordinate at about $\omega = .53$: as the cutoff increases we get higher estimates. To get an idea of the level of smoothing implied by the local likelihood estimate using $\omega_c = \pi/4$, one should refer to the central panels of figure 7. For the U.S. $\hat{\theta}$ is negative and high for low cutoffs, and increases more gradually than in the Italian case.

[Figure 9 about here]

8 Conclusions

With respect to two well-known and widely used models, the local level and the local linear trend models, this paper has evaluated estimation strategies alternative to maximum likelihood, namely multistep (adaptive) estimation, cross-validation and local likelihood in the frequency domain, for the purposes of long-range forecasting and the decomposition into a trend and a cyclical component from economic time series.

After introducing a metric for assessing the performance of the misspecified model, it has been shown that, although forecast efficiency is not the same as efficiency in the estimation of unobserved components, multistep estimation is an effective strategy for the purposes considered: the examples in the paper show illustrate that the signal extraction efficiency can be very high.

Secondly, cross-validatory estimates tend to be very close to those minimising the variance of the one-step ahead prediction errors, and thus to maximum likelihood; loosely speaking, cross-validation gives more weight to the high frequency components in the series, compared to multistep estimation. The resulting estimates (that optimise the leave-one-out interpolation performance of the approximating model) tend to be of little value for the extraction of signals from a time series. We leave to future research the assessment of multiple cross-validation, which can be performed using the algorithm proposed in Proietti (2003).

Finally, local likelihood provides an alternative way of conceptualising and operationalising the notion of constructing (possibly simple) predictors and signal extraction filters that emphasise the long run features of a series.

References

Bhansali, R.J. (2002), Multi-step forecasting, in Clements, M.P. and D. F. Hendry (eds.), *A Companion to Economic Forecasting*, Blackwell Publishers, Oxford, 206-221.

Clements, M.P. and Hendry, D.F. (1996). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics*, 58, 657-684.

Cox, D.R. (1961). Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society*, Series B, 23, 414-422.

de Jong, P. (1988). A cross-validation filter for time series models. *Biometrika*, 75 , 594-600.

Doornik, J.A. (2001), *Ox. An Object-Oriented Matrix Programming Language*, Timberlake Consultants Press, London.

Green, P.J. and Silverman, B.W. (1994), *Nonparametric regression and generalized linear models*, Chapman & Hall, London.

Gómez, V. (2001). The Use of Butterworth Filters for Trend and Cycle Estimation in Economic Time Series. *Journal of Business and Economic Statistics*, Vol. 19, No 3, 365-373.

Hannan, E.J., and Nicholls, D.F. (1977). The estimation of the prediction error variance. *Journal of the American Statistical Association*, 72, 834-840.

Harvey, A.C. (1989), *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge University Press, Cambridge, UK.

Haywood, G. and Tunnicliffe Wilson, G. (1997), Fitting Time Series Models by Minimizing Multistep-ahead Errors: a Frequency Domain Approach, *Journal of the Royal Statistical Society, Series B*, 59, 237-254.

Hodrick R.J., and Prescott, E.C. (1997), Postwar U.S. Business Cycles: an Empirical Investigation, *Journal of Money, Credit and Banking*, 29, 1-16.

Kohn, R., and Ansley, C.F. (1989). A fast algorithm for signal extraction, influence and cross-validation in state space models. *Biometrika*, 76(1), 65-79.

Kohn, R., Ansley, C.F. and Tharm, D. (1991). The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *Journal of the American Statistical Association*, 86, 1042-1050.

Kohn, R., Ansley, C.F. and Wong, C.H. (1992). Nonparametric spline regression with autoregressive moving average errors. *Biometrika*, 79, 335-46.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika*, 80, 117-26.

Koopman S.J., Shepard, N., and Doornik, J.A. (1999), Statistical algorithms for models in state space using SsfPack 2.2, *Econometrics Journal*, 2, 113-166.

Muth, J.F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, 55, 299-306.

Pedregal D. J. and Young P. (2001). Some comments on the use and abuse of the Hodrick-Prescott filter. *Review on Economic Cycles, International Association of Economic Cycles*, 2001, vol. 2.

Proietti, T. (2002), Forecasting With Structural Time Series Models, in Clements, M.P. and D. F. Hendry (eds.), *A Companion to Economic Forecasting*, Blackwell Publishers, Oxford, 105-132.

Proietti T. (2003). Leave-k-out diagnostics for state space models. *Journal of Time Series Analysis*, 24, 221-236.

Proietti T., and Harvey, A.C. (2000). A Beveridge Nelson Smoother. *Economics Letters*, 67, 139-146.

Tiao, G. C., and Xu, D. (1993). Robustness of Maximum Likelihood Estimates for Multi-Step Predictions: The Exponential Smoothing Case. *Biometrika*, 80, 623-641.

Tiao, G. C., and Tsay, R. S. (1994), "Some advances in non-linear and adaptive modelling in time-series", *Journal of Forecasting*, 13, 109-131.

West, M. and P.J.Harrison (1997). *Bayesian Forecasting and Dynamic Models.* New York: Springer-Verlag.

Whittle P. (1983) *Prediction and Regulation by Linear Least Squares Methods*, Second edition. Basil Blackwell, Oxford.

Young P. and Pedregal D. (1999), Recursive and en-bloc approached to signal extraction. *Journal of Applied Statistics*, 26, 103-128.

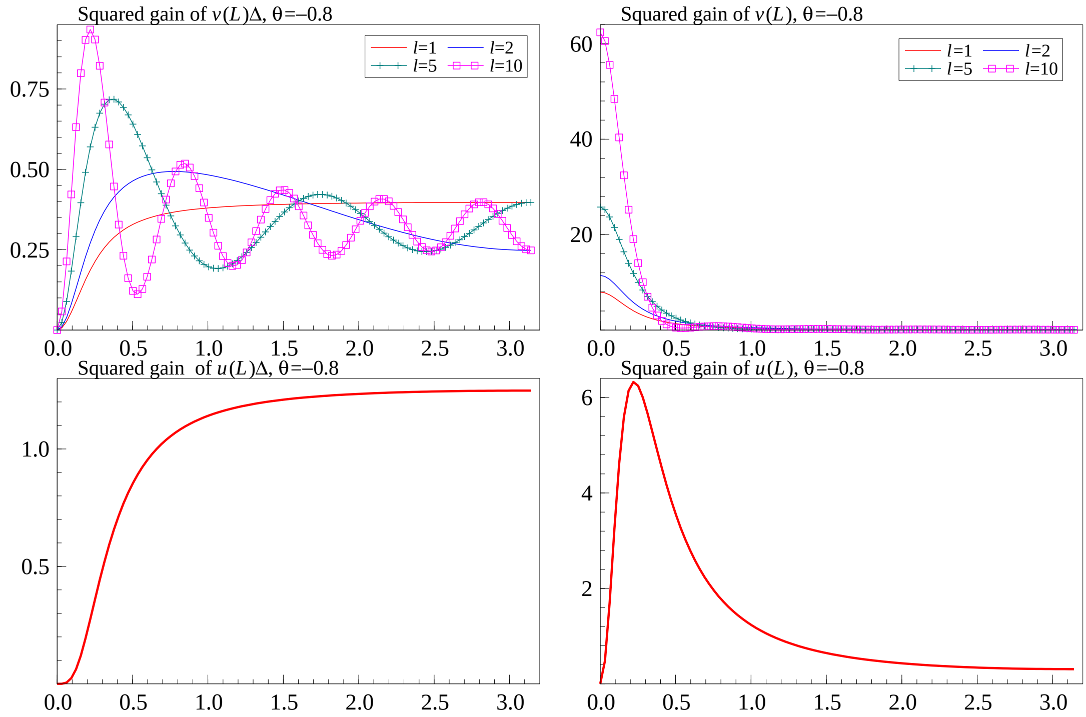

Figure 1: Local linear model: squared gains of ME and CV filters.

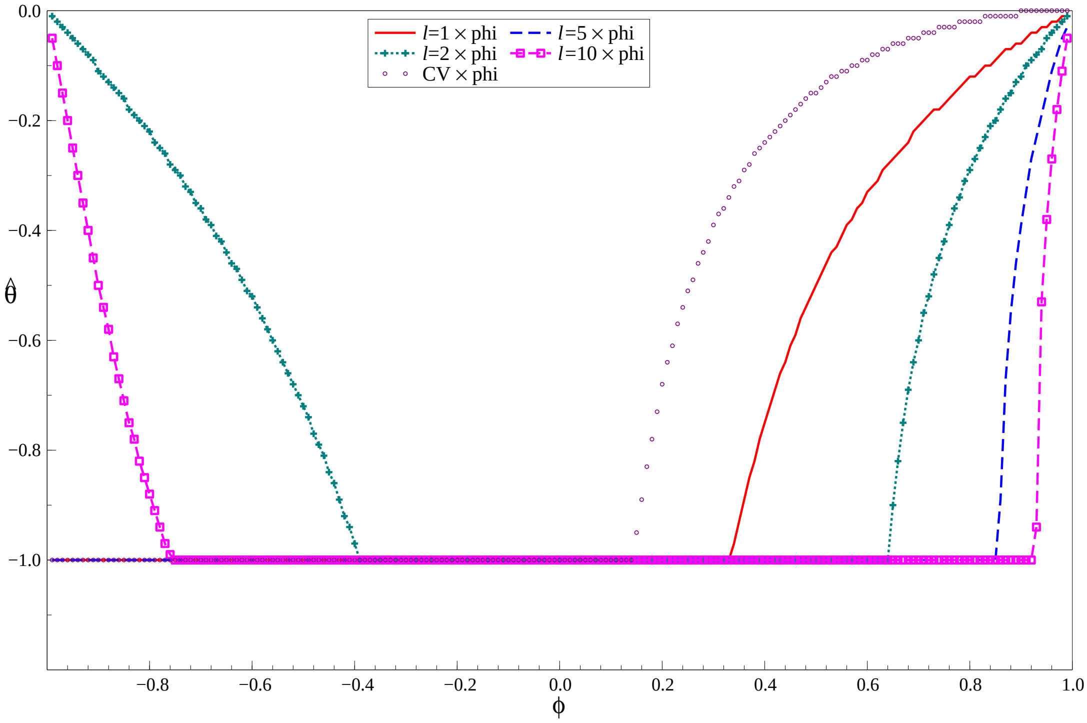

Figure 2: ME and CV estimates of the LLM parameter θ for the AR(1) model $y_t = \phi y_{t-1} + \xi_t$.

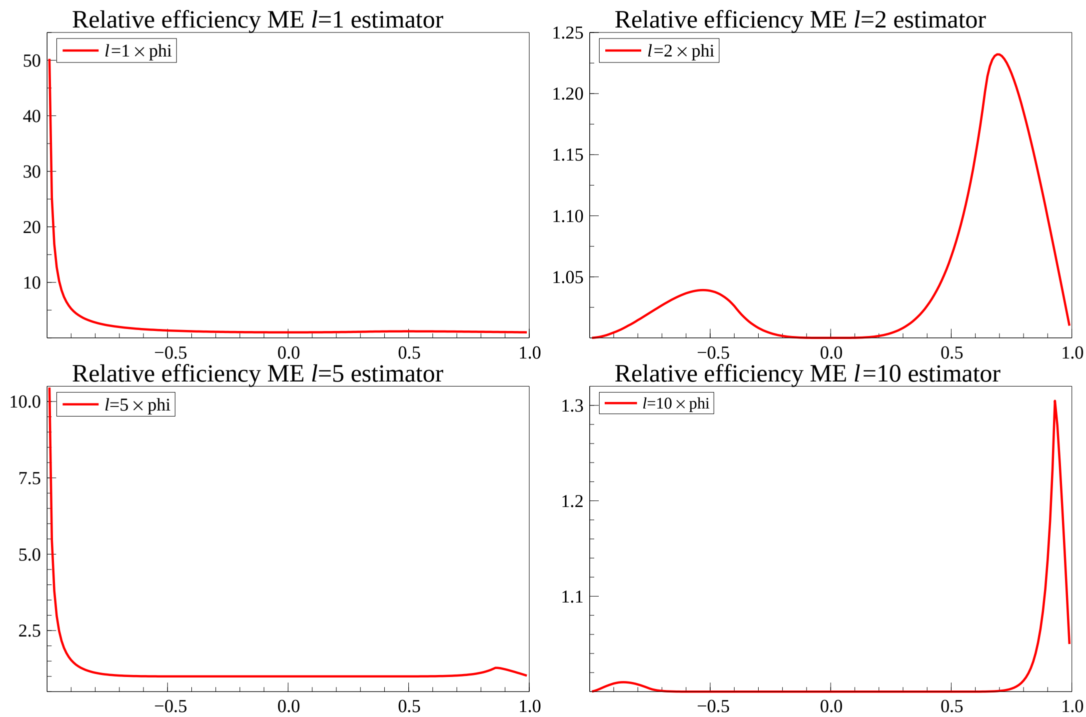

Figure 3: Relative forecast efficiency of ME(l) estimates of the LLM parameter θ for the AR(1) model $y_t = \phi y_{t-1} + \xi_t$.

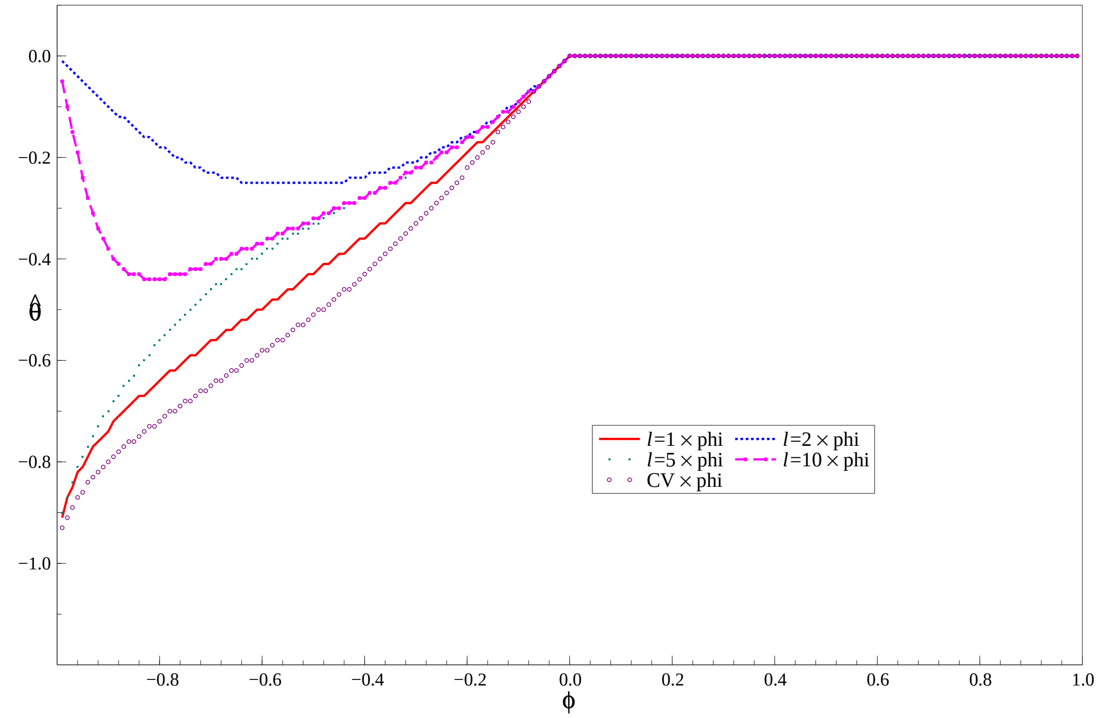

Figure 4: ME and CV estimates of the LLM parameter θ for the ARIMA(1,1,0) model $\Delta y_t = \phi \Delta y_{t-1} + \xi_t$.

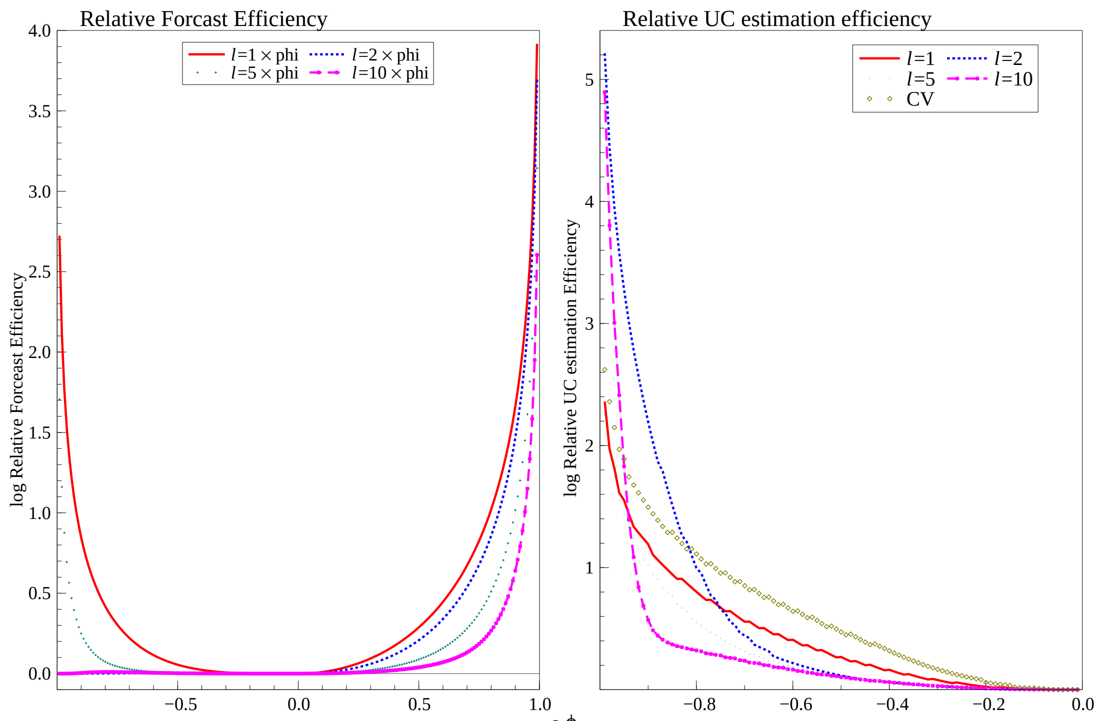

Figure 5: Relative forecast (left panel) and UC estimation efficiency (right panel) of ME(l) estimates of the LLM parameter θ for the ARIMA(1,1,0) model $\Delta y_t = \phi \Delta y_{t-1} + \xi_t$.

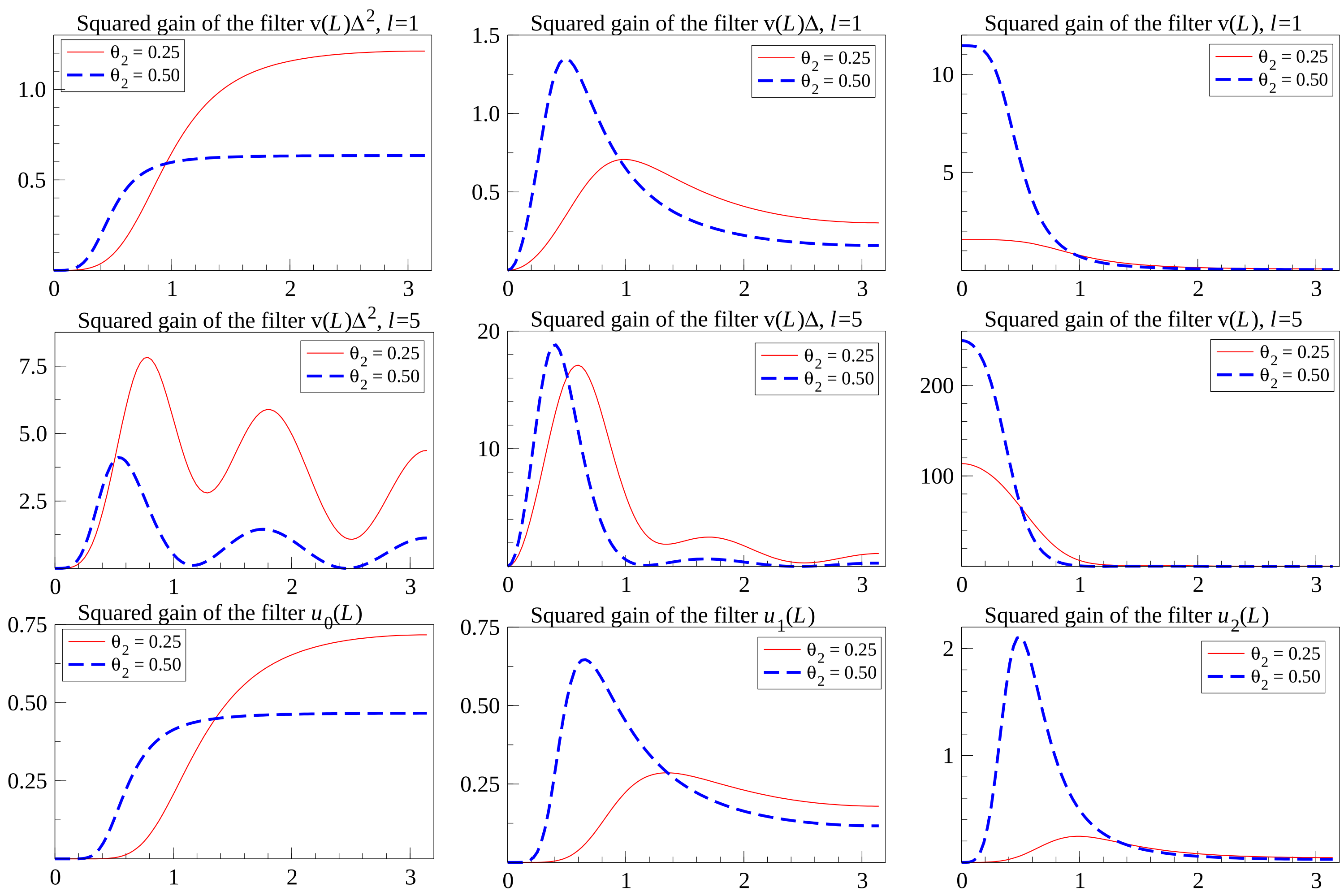

Figure 6: Local linear trend model: squared gains of ME and CV filters.

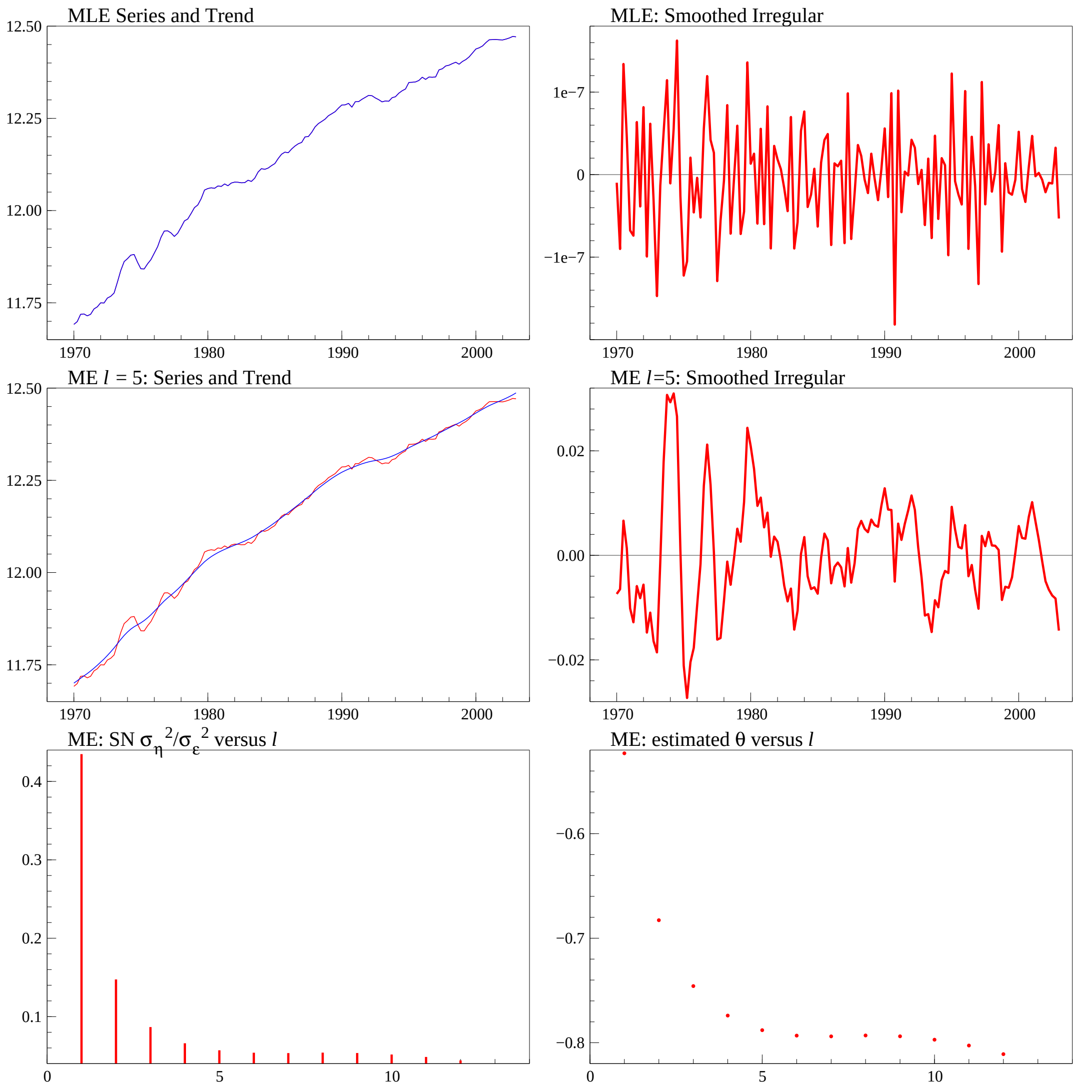

Figure 7: Trend and cycle in Italian real GDP at constant prices extracted by a local linear model with constant drift using maximum likelihood estimation (top panels) and multistep estimation with $l \leq 12$. The estimates of the signal to noise ratio and of the MA parameter θ of the IMA(1,1) reduced form display an elbow at $l = 5$.

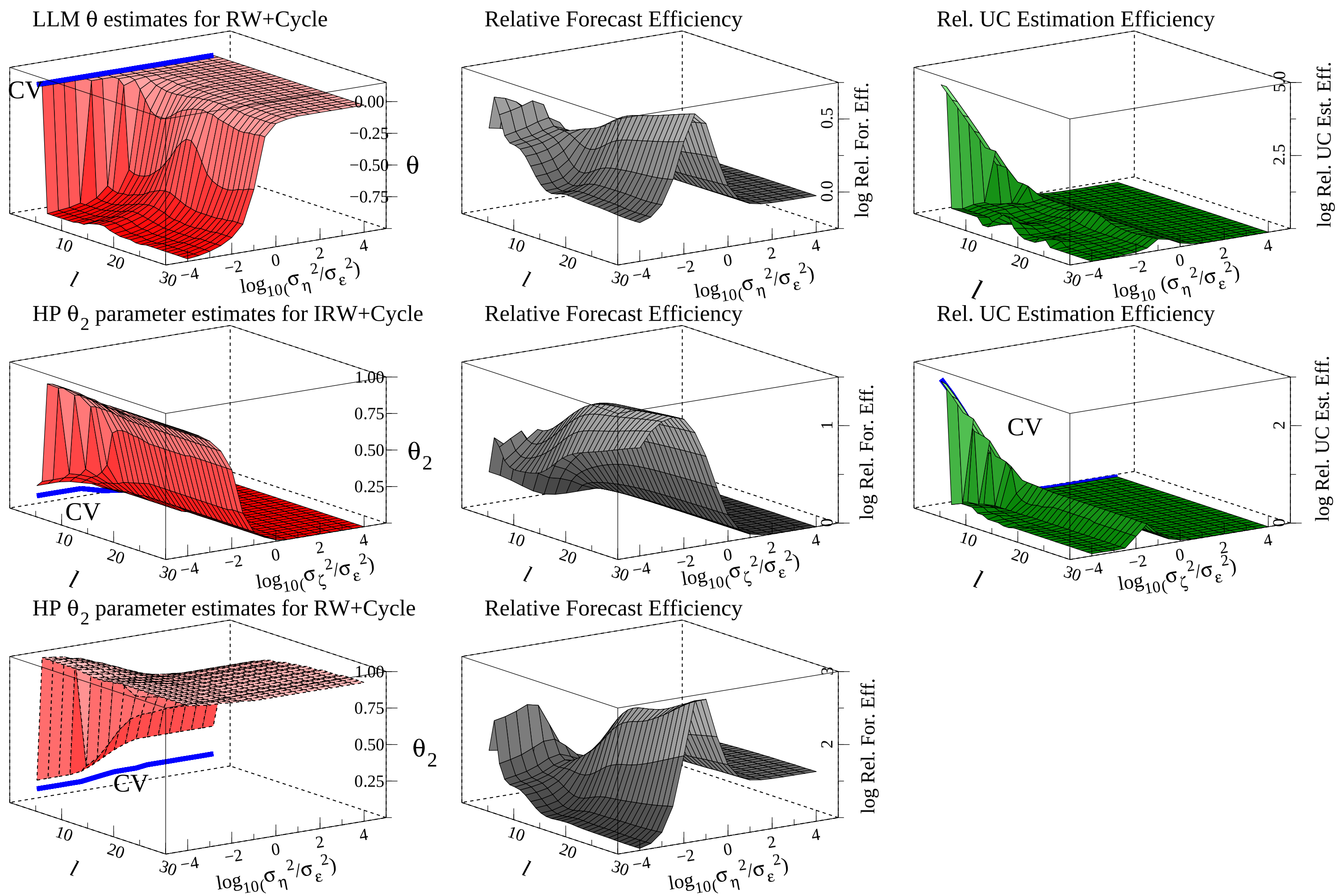

Figure 8: ME and CV estimates of the LLM and HP models and relative efficiency.

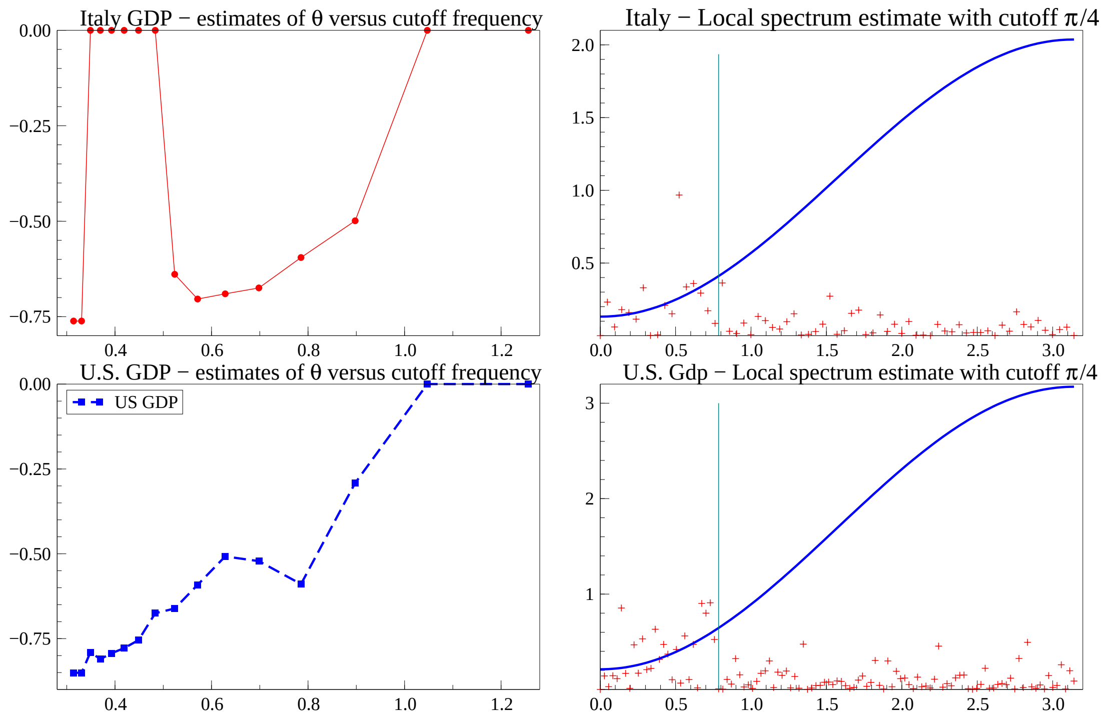

Figure 9: Local likelihood estimation of the local level model for the Italian and U.S. quarterly GDP (logarithms). The plots on the left hand side display the estimated θ values for cutoff frequencies in the range $\pi/10$ (corresponding to a period of 5 years) and $2\pi/5$ (corresponding to a period of 5 quarters). Those on the right hand side display the periodogram of Δy_t, and the parametric spectral fit using the periodogram up to the cutoff frequency $\pi/4$ (corresponding to a period of 2 years).