

Literature review of modern time series forecasting methods

(This document covers the stochastic linear model approaches)

By Paul Karapanagiotidis

July 31, 2012

Modern time series forecasting methods are essentially rooted in the idea that the past tells us something about the future. Of course, the question of how exactly we are to go about interpreting the information encoded in past events, and furthermore, how we are to extrapolate future events based on this information, constitute the main subject matter of time series analysis.

Typically, the approach to forecasting time series is to first specify *a model*, although this need not be so. This model is a statistical formulation of the dynamic relationships between that which we *observe* (i.e. the so called *information set*), and those variables we believe are related to that which we observe. It should thus be stated immediately that this discussion will be restricted in scope to those models which can be formulated *parametrically*.

The “classical” approach to time series forecasting derives from regression analysis. The standard regression model involves specifying a linear parametric relationship between a set of *explanatory variables* (or *exogenous* variables) and the *dependent* (or *endogenous* variable). The parameters of the model can be estimated in a variety of ways, going back as far as Gauss in 1794 with the “Least Squares” method, but the approach always culminates in striving for some form of statistical orthogonality between the explanatory variables and the residuals (or *innovations*) of the regression. That is, we wish to express the linear relationship in a dichotomous form in which the *innovations* represent that part of our information which is completely unpredictable. It should probably also be emphasized that in the engineering context this is analogous to reducing a signal to “white noise.”

However, this review is to be concerned with more “modern” approaches and in many ways, it was the practical necessities of engineering that provided an initial impetus. Both Wiener (1949) and Kolmogorov (1941) were pioneers in the field of linear prediction, and while their approaches differed (Wiener worked in the frequency domain popular amongst engineers, while Kolmogorov worked in the time domain), it is clear that their solutions to the same basic geometrical problem were equivalent (see Priestley (1981) ch.10). Wiener’s work, in particular, was especially relevant to modern time series forecasting in that he was among the first to rigorously formulate the problem of “signal extraction.” That is, given observations on a time series corrupted by additive noise, what is the optimal estimator (in the mean-squared error (MSE) sense) of the latent or underlying signal (or *state* variable).

Given the historical context of massive systems of equations models popular among macroeconomic forecasters of 1950’s (see for example the Klein-Goldberger model (1955) or Adelmans (1959) for details) it quickly became apparent that forecasting models derived from a signal

extraction context forecasted at least as well as those based on complicated systems of economic relationship equations formulated as individual, yet interconnected, dynamic classical regressions.

Methods such as the U.S. Bureau of the Census X-11 and its later developments such as the X-11 ARIMA and X-12 ARIMA¹ were able to extract seasonal and other cyclical component signals from a series by means of a iterated finite moving average procedure. Later developments improved on this general method. For example, the simple exponentially weighted moving average (EWMA)² allowed for geometrically decreasing, infinite, causal, impulse response. Holt (1957) and Winters (1960) further generalized this method to include a slope component in the forecast function (since the EWMA implies a horizontal forecast function) and also allowed for seasonal effects. Later, Brown (1963) reformulated the problem in terms of a discounted least squares regression, termed “Linear exponential smoothing.” It turns out that Brown (1963) is a special case of Holt and Winters (1957,1960).

These methods became popular due to their simplicity and ease of application for the general practitioner. However, these methods are considered *ad hoc* since they fail to incorporate theoretical considerations into their decompositions of cyclical components and they are formulated without recourse to a well specified statistical model. Consequently, they also do not allow for prediction intervals since there is no accounting for predictive variance.

Despite this, Makridakis and Hibon (1979) seemed to suggest that simple *ad hoc* methods tended to forecast at least as well as more complicated methods. However, in response Newbold (1983) contended that it is not with slavishly applying different *methods* that we should be concerned with but rather, that we should consider the relevant issues involved in forecasting *any particular series in its own right*. That is, the choice of *method* should be of secondary importance to the *approach* – which itself depends on characteristics of the particular series—and that our time would be better spent in considering case studies which evaluate how various *approaches* to forecasting complement the nature of the data, as opposed to “blind” horseraces between *methods*.

With this in mind, it seems important then to consider more modern approaches which permit us greater flexibility in specifying a well-defined statistical model, as opposed to restrictive *ad hoc* methods. Moreover, this approach should free us to incorporate an *economic rationale* for our choice of particular forecasting method. It is in this sense that perhaps the first great example of this type of analysis was presented by Muth during the early 1960’s.

In 1960, Muth established the statistical properties of the EWMA in terms of the Wold decomposition. It was shown that this method of signal extraction implied: 1) a constant forecast function across forecast horizons 2) a random walk plus noise model in the levels of the series. The former comes about since the forecast is a weighted average of the past innovations but where the weights are all equal; that is, “the best forecast for the time period immediately ahead is the best forecast for *any* future time period, because both give estimates of the ‘permanent’ component.” The

¹ see Findley et. al. (1992,1996) for an overview, or Macaulay (1931) for the original exposition.

² This is also referred to as “exponential smoothing” in the literature.

latter is derived directly from the fact that $E[y_t | \varepsilon_{t-1}, \dots] = \sum_{i=1}^{\infty} \omega_i \varepsilon_{t-i} = y_t^e = \sum_{i=1}^{\infty} v_i y_{t-i}$ implies, under $\min E(y_t - y_t^e)^2$ and $y_t = \sum_{i=1}^t \varepsilon_t + \eta_t$ s. t. $\varepsilon_t, \eta_t \sim WWN$ that $v_i = \beta(1 - \beta)^{i-1}$ s. t. $\beta \in (0,1)$.

Muth (1960) was motivated by concerns he had over the notion of *Adaptive Expectations* as formulated by Cagan (1956) and applied by Friedman (1957) within the context of Friedman's Permanent Income hypothesis. Since Adaptive Expectations can be formulated as an EWMA on the levels of the series, Muth's problem (as illustrated above) was to find the "inverse optimal predictor"—that is, for what discrete-time, univariate, stochastic process is the discrete-time version of the adaptive expectations mechanism (i.e. EWMA) optimal in the MSE sense? (Hansen and Sargent, 1983). The answer was of course, a random walk plus noise process in the levels—one which held rather strict implications for economic time series forecasting.

In 1961, Muth further generalized the statistical formulation of Adaptive Expectations into what he called Rational Expectations. This new conception of expectations formulation implied a more general form for the stochastic linear representation of the series than the random walk plus noise model, and represented a major precursor to the later "structural time series" approach of Nerlove, Harvey, Young and others.³ Moreover, this generalization allowed us more flexibility to incorporate various assumptions about the underlying DGP of the levels process according to the prescriptions of economic theory.

Nerlove (1967), building on Muth's ideas and his own past work on optimal lag distributions⁴ (see Nerlove and Wage (1964) for example), heavily emphasized theoretical contribution to model selection specifically that of appropriately specified expectations formation. Here, rather than assume that agents respond to forecasts of future values of economic variables they directly observe, Nerlove defined what he called "unobserved components" (UC) or underlying economic variables *not directly observable* by agents; postulating instead that agents respond to the minimum mean squared error "extractions" of these unobserved components.⁵ Furthermore, Nerlove also emphasized that the nature of the selected decomposition should be derived from an optimization model over time (although it was not until Taylor (1970) that this was done, within the context of production theory of the firm).

Despite the valuable contributions to frequency-domain analysis that Nerlove and others made to the study of seasonal and other cyclical features of time series during the late 1960's (see for example Nerlove and Grether (1970)), the 1970's were to become dominated by the time-domain analysis techniques advocated by Box and Jenkins (1970). There are various reasons for this. The main reason perhaps was that Box and Jenkins provided a complete methodology that resolved many practical issues in applying time series forecasting and optimal control, and did so in a way that was easy for the analyst to implement. For example, Box and Jenkins focused on the so called *transfer function models* which

³ "Structural" in this sense differs from the classical systems of equations interpretation. Here what is meant is that some sort of stochastic structure is imposed on the levels of the series y_t so as to capture cyclical components within a decomposition. That is, all "structural" models are the direct stochastic analog of $y_t = a + bt + \sum_j c_j z_{j,t} + \varepsilon_t$ where $\varepsilon_t \sim WWN$, t is a linear deterministic time trend, and the $z_{j,t}$'s are seasonal dummy variables.

⁴ By "lag distribution" in this context what is meant is the autoregressive coefficient function.

⁵ It is worth noting that nearly all of his formal presentation draws directly from Wiener (1949).

provided an easily estimable time-domain representation of a frequency-domain concept familiar to engineers and system control theorists. The notion that an infinite, causal, impulse response function might well be approximated by a ratio of finite z-polynomials was an important step. Second, Box and Jenkins provided a way around the problem of non-stationarity by means of a methodology focused on *differencing* the data. At this point, Bell (1984) had not yet formulated the solution to the Wiener optimal signal plus noise extraction problem under non-stationarity. And even if he had, the computational capabilities of the time were unable to handle the complex Fourier operations required to practically implement a Wiener type frequency-domain analysis in any practical way.⁶ Finally, it was suggested that ARIMA models might well make suitable forecasting models since it turned out that many of the previous *ad hoc* methods were represented as special cases: for example the ARIMA(0,1,1) model leads to forecasts produced by an EWMA, while the ARIMA(0,2,2) corresponds to Holt's method (1957).⁷

Moreover, Box and Jenkins also provided a simple means of modeling seasonality within the time-domain. The so called "airline model"⁸ proposes a rational impulse response function of the type $\varphi(L) = \frac{(1+\varphi L)(1+\varphi L^s)}{\Delta \Delta_s}$ so that $y_t = \varphi(L)\xi_t$ s.t. $\xi_t \sim WWN$. The success of the airline model led to the general multiplicative seasonal SARIMA process of order $(p, d, q) \times (P, D, Q)_s$: $\Phi(L^s)\varphi(L)\Delta^d\Delta_s^D y_t = \theta_0 + \Theta(L^s)\theta(L)\xi_t$. Therefore the airline model is the special case of $(0,1,1) \times (0,1,1)_s$ with no constant.

However, it should be noted that despite the success of the Box-Jenkins approach to forecasting during the 1970's, there were still some who chose to work within a structural time series framework. For example, Harrison and Stevens (1976) were successful in formulating the linear, Gaussian, Markovian state-space model within a Bayesian context. Working with the Kalman filter, they were able to specify their "dynamic linear model" (DLM) based on time varying parameters in order to account for nonstationarity. Moreover, the Bayesian approach allows one to specify prior distributions on not only parameters, but also the initial conditions, facilitating convergence of the Kalman gain matrix. Furthermore, Harrison and Stevens is important in another respect since it represented an important foundation for the later non-linear *regime switching* models of the late 1980's (see for example Hamilton, 1989).

During the 1970's Nerlove continued working within the UC framework, culminating in his book with Grether and Carvalho (Nerlove et. al. 1979). This volume represented the cornerstone of future UC

⁶ It is of note in fact that two other developments arose specifically in an attempt to circumvent the computational problem: the development of the Kalman filter in (1960) and the Cooley-Tukey Fast Fourier Transform algorithm in (1965). Furthermore, in Schweppe (1965) it was demonstrated that the likelihood function of the Gaussian linear model employed in the Kalman filter could be evaluated via the prediction error decomposition which made parameter estimation feasible. However, despite the usefulness of these novel approaches, neither was widely applied within the econometrics literature at least until the late 1970's (see for example Engle (1978) or Harrison and Stevens (1976)). During the early 1980's, however, Harvey widely advocated use of the Kalman filter to estimate his generalization of the Unobserved Components models established by Nerlove in (1967).

⁷ See Box and Jenkins (1970) pg.105-112.

⁸ See Box and Jenkins (1970) ch.9

modeling, including work by Harvey during the 1980's. While not only containing an illuminating history of UC as an epistemological principle all the way back to the days of Kepler, the authors laid out in great detail the probabilistic foundations, formulation, analysis, and estimation of UC models. Chapter 4 deals with the so called *canonical form* of the UC model⁹ and provides a general algorithm for deriving it given more complicated UC formulations. Note that typically the particular specification of the structural form will imply restrictions on the parameters of the *canonical form* (i.e. reduced form).¹⁰ Chapter 6 also discusses how to go about choosing an appropriate UC form – note that as opposed to the Box and Jenkins approach of appealing to the time-domain features through the sample (partial) autocorrelation function (SACF), Nerlove suggests rather that the features of the UC decomposition should be informed by the *spectral characteristics* of the time series. In fact, one of the major criticisms facing ARIMA modeling of the Box-Jenkins type was the over reliance on the SACF estimator (within the so called “identification stage”) which was known to exhibit poor statistical properties (see Hannan (1960)). Estimation methods are provided in both the time-domain (by maximum-likelihood estimation of the parameters of the reduced form ARMA model implied by the *canonical form* and then mapping these estimates back into the parameters of the UC form) and in the frequency domain (avoids the need deriving the *canonical form* entirely) by a method suggested by Hannan (1969).

Engle (1978) continued to work within the UC framework, specifically dealing with estimating structural models of seasonality in response to work done by Nerlove and Grether (1970), Nerlove (1967) and Nerlove (1972). Engle defined what he calls the *UCARIMA model* (unobserved components autoregressive moving average), suggesting the class of ARIMA processes as ideal components within a UC framework intended for modeling seasonality. The UCARIMA representation decomposes the process $y_t = T_t + S_t + C_t + I_t$ into individual ARIMA processes (where T_t is the trend component, S_t is some seasonal component, C_t is a non-seasonal cyclical component, and I_t is some residual irregular component.) For example Harvey's *Local Level (with random walk)* model is actually a special case of the UCARIMA framework where only the trend and irregular components are modeled together as a single ARIMA process. The distinguishing feature of Engle's UCARIMA framework, however, is not that it decomposes the process—it is rather that the components are explicitly defined as a *statistical model* of seasonal variation, as opposed to purely *ad hoc* signal extractions methods like the X-11. Moreover, in the vein of Pagan (1975), Engle employs the state-space framework and Kalman filter to estimate the latent components themselves once the parameters of the structural form are provided (Engle assumes they are known).

Finally, it is interesting to note that this paper highlights some of the difficulties UC models present, such as the often bewildering identification conditions, especially in the multivariate case. In this sense, while the structural form implies a specific UCARIMA form, the UCARIMA form corresponds with an infinite number of possible structural forms. Therefore the class of structural models forms a subset of the UCARIMA models and it is therefore possible that there does not exist a structural form for all

⁹ That is, the reduced form $y_t = \frac{G(z)}{H(z)} \varepsilon_t$ where y_t has been rendered stationary by some method, perhaps differencing, wherein the differencing operators would show up as factors of $H(z)$.

¹⁰ See Harvey (1989) pg.67 for some simple examples given basic structural UC forms.

UCARIMA specifications (Ledolter (1984) cites this as a general deficiency of the structural approach, as the same argument also applies to Nerlove's *canonical form* representation).

Despite these issues, during the 1980's authors such as Harvey (1984) brought the use of the UC framework, for both modeling and forecasting economic time series, into the mainstream. Harvey (1984), entitled "A unified view of statistical forecasting procedures," set out to advocate that a UC, or *structural* framework, where the structure of a time series is decomposed into trend, seasonal and irregular components, could both account for and generalize previous approaches such as the *ad hoc*, Box-Jenkins, and Harrison-Stevens methodologies. Later, in 1989, Harvey also consolidated many of the results of this period into a book "Forecasting, structural time series models, and the Kalman filter." However, it is not so clear what Harvey specifically contributed to the UC framework that wasn't already achieved by Nerlove or Harrison and Stevens. For example, it was J. Ledolter (1984) that suggested in his comments of Harvey (1984) that "apart from a somewhat different and also more restrictive characterization of the seasonal component, this model is basically the same as the one developed by Harrison and Stevens (1976)."

Ultimately, Harvey's main contributions lay in the area of estimation theory. For example, in Harvey and Franzini (1983) the authors develop a test between the null hypothesis of a deterministic trend plus seasonal model of the type $y_t = \alpha + \beta t + \sum_j \delta_j z_{j,t} + \varepsilon_t$,¹¹ and the alternative stochastic analog model (the *Local Linear Trend*) $y_t = \mu_t + \gamma_t + \varepsilon_t$ s.t. $\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$, $\beta_{t-1} = \beta_{t-2} + \xi_{t-1}$ where μ_t represents the trend ($\eta_t, \xi_t \sim WWN$) and $\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t$ defines the seasonal component ($\omega_t \sim WWN$).

It was also suggested within Harvey and Todd (1983) that *Local Linear Trend* models of the kind above provide a useful alternative to the seasonal SARIMA models of Box-Jenkins. That is, instead of working first with some SARIMA model and imposing restrictions so as to ensure the existence of a structural decomposition (i.e. trend, seasonal, and irregular) (as is done in Hillmer and Tiao (1982) in a way consistent with an structural model estimated by X-11), the authors work in the opposite direction, specifying first the structural form which will then generate the required reduced form parameter restrictions directly.¹² The structural approach is then set in a horserace against the Box-Jenkins seasonal forecasting approach and it is suggested that the structural approach compares well with the alternative.

Harvey was also an early adopter of the application of the Kalman filter to the problem of time series estimation. For example, in Harvey and Phillips (1979) the authors cast a regression model with ARMA disturbances in a state-space form and employ the Kalman filter in avoiding computational difficulties associated with MA type disturbances and inversion of their associated covariance matrix. Moreover in 1980, Harvey, Gardner, and Phillips published an algorithm, along with Fortran 66 code, for the exact Maximum Likelihood estimation of ARMA models by means of Kalman filtering. Later, Harvey

¹¹ Where $\varepsilon_t \sim WWN$, t is a linear deterministic time trend, and the $z_{j,t}$'s are seasonal dummy variables.

¹² See again the Nerlove *canonical form* discussion above.

applied the Kalman filter to the extraction of unobserved components (see Harvey and Todd (1983), and Harvey (1984),(1989), and Harvey and Jaeger (1993)).

Note that trend models of the *Local Linear Trend* type above represent what Young (2011) ch4. pg.66 refers to as the *Generalized Random Walk* (GRW), since they include Harvey’s *Local Linear Trend* model (see Harvey (1984,89) or Harrison and Stevens (1976)),¹³ the *Integrated Random Walk*, the *Smoothed Random Walk*, the AR(1) process, and *Damped Trend* all as special cases. The flexibility of this type of stochastic trend model is significant since it avoids the issue of spurious relationships implied by deterministic trends of the type $\mu_t = \mu + \beta t$ (see Nelson and Kang (1981,84)).¹⁴ Note that the *Local Linear Trend* model is significant in the sense that it generalizes the *ad hoc* method proposed by Holt (1957) to the signal plus noise case. Take for example that Holt’s recursions can be re-expressed as:

$$m_t = \lambda_0 y_t + (1 - \lambda_0)(m_{t-1} + b_{t-1}) = m_{t-1} + b_{t-1} + \lambda_0 v_t$$

$$b_t = \lambda_1 (m_t - m_{t-1}) + (1 - \lambda_1)b_{t-1} = b_{t-1} + \lambda_0 \lambda_1 v_t$$

which is the same form of the *Local Linear Trend* model, except the innovations are not stochastic i.e. $v_t = y_t - m_{t-1} - b_{t-1} = y_t - y_{t|t-1}$. Therefore, the signal is *observable* in Holt’s filter; that is we assume a (m_0, b_0) and then each (m_t, b_t) are known with certainty given the filter parameters (λ_0, λ_1) and the data y_t . This contrasts with the *Local Linear Trend* model where the signal for the trend μ_t and slope β_t must be *extracted* from the additive noise (ξ_{t-1}, η_t) and the observations y_t . This is exactly why the Kalman filter is useful within this context, since the *Local Linear Trend* model is by construction a state-space formulation. Moreover, note that under the *Local Linear Trend* model, we must specify an explicit statistical model and in doing so we impose *probability laws* on the stochastic process for the trend. It is these laws that then allow us to do *forecast inference* which is not possible with the *ad hoc* filtering methods. Finally, the nature of a slow moving stochastic trend is its ability to parsimoniously capture and extrapolate what might otherwise possibly be an extremely non-linear (maybe even chaotic) “true” underlying deterministic process that exists in the real world.

In contrast to the UC approach of Harvey, Nerlove, and others, Beveridge and Nelson (1981) described an alternative form of decomposition. As the paper’s introduction explains, since Nerlove (1967) had approached the problem of an appropriate time series decomposition from a signal extraction perspective, he faced the fundamental problem of prior knowledge of the particular stochastic process generating his unobserved components.¹⁵ Moreover, it was suggested that studies such as Mintz (1969) that attempted to isolate business cycles using *ad hoc* methods such as the X-11 were flawed both practically and theoretically—the former because of problems with extrapolation¹⁶ inherent in the X-11 method and theoretically because “ideally the procedure for extraction of trend

¹³ Note that the *Local Linear Trend* model is actually due to Harrison and Stevens (1976). See page 217.

¹⁴ It is worth noting that this type of deterministic trend function formed the basis of Brown’s (1963) discounted least squares *Linear Exponential Smoothing* (LES) filter. Furthermore, note that if $\beta = 0$ then LES reduces to EWMA (see Harvey (1984) pg.252).

¹⁵ Which incidentally was also a motivation behind the Bayesian approach employed by Harrison and Stevens (1976).

¹⁶ The so called “end-points” problem.

should be appropriately tailored to the stochastic properties of each series considered.” Furthermore, the authors also comment on Muth (1960) in suggesting that while “an important virtue of [Muth’s] approach is its freedom from determinism ... to interpret ‘permanent’ as ‘expected’ requires rather strong prior assumptions about the stochastic structure of the series in question.” It is easy to counter these statements, however. First, the fundamental problem of prior knowledge a signal extraction approach entails can also be viewed more positively as a strength of the framework, not a limitation, since it allows us to both build and investigate models where the trend is variously specified under different possible stochastic formulations. In fact, the Beveridge-Nelson (BN) decomposition suffers from a very rigid interpretation of trend and cycle: that is, the BN decomposition implies a pure random walk with drift trend where cycles are simply differences of the current level of the series from the trend. Therefore, it seems strange that BN would criticize Muth’s analysis on the grounds of generalizability since that wasn’t even the point of Muth (1960) to begin with ¹⁷ and the BN decomposition isn’t quite so general itself. In fact, Zivot et. al. (2007) show that there “always exists a UC representation with perfectly correlated shocks that is consistent with the BN decomposition.” Moreover, this representation is not unique.

¹⁷ The point of Muth (1960) was to demonstrate the issues inherent in an *Adaptive Expectations* formulation and to lay groundwork for the generalization outlined in Muth (1961).

Bibliography

- I. Adelman, L. Adelman (1959) "The dynamic properties of the Klein-Goldberger model," *Econometrica*, 27(4) 596-625
- W. Bell (1984) "Signal extraction for nonstationary time series," *The Annals of Statistics*, 12(2), 646-664
- G. Box and G. Jenkins (1970) *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.
- R.G. Brown (1963) *Smoothing, Forecasting and Prediction*, Englewood Cliffs: Prentice Hall
- P. Cagan (1956) "The monetary dynamics of hyper-inflation," pp 25-117 in M. Friedman, ed, *Studies in the Quantity Theory of Money*, Chicago: University of Chicago Press
- J.W. Cooley, J.W. Tukey (1965). "An algorithm for the machine calculation of complex Fourier series". *Math. Comput.* 19 (90), 297–301.
- R. Engle (1978) "Estimating structural models of seasonality," in *Seasonal Analysis of Economic Time Series*, A. Zellner ed., NBER
- D.F. Findley, B.C. Monsell, M.C. Otto, W.R. Bell, M. Pugh (1992) "Towards X-12 ARIMA," Technical report, Bureau of the Census
- D.F. Findley, B.C. Monsell, W.R. Bell, M.C. Otto, B.C. Chen (1996) "New capabilities and methods of the X-12 ARIMA seasonal adjustment program," Technical report, U.S. Bureau of the Census
- M. Friedman (1957) *A Theory of the Consumption Function*, New Jersey: Princeton University Press
- J.D. Hamilton (1989) "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57(2), 257-384
- E.J. Hannan (1960) *Time Series Analysis*, London: Methuen
- _____ (1969) "The estimation of mixed moving average autoregressive systems," *Biometrika*, 56, 579-593
- Hansen and Sargent (1983) "Aggregation over time and the inverse optimal predictor problem...", *Internal Economic Review*, 24
- P.J. Harrison, C.F. Stevens (1976) "Bayesian Forecasting," *Journal of the Royal Statistical Society Series B*, 38(3), 205-247
- A.C. Harvey (1984) "A unified view of statistical forecasting procedures," *Journal of Forecasting*, 3(3), 245-275
- _____ (1989) *Forecasting, structural time series models, and the Kalman filter*, Cambridge University Press.

- A.C. Harvey, L. Franzini (1983) "Testing for deterministic trend and seasonal components in time series models," *Biometrika*, 70, 673-682
- A.C. Harvey, G. Gardner, G. Phillips (1980) "An algorithm for exact maximum likelihood estimation by means of Kalman filtering," *Applied Statistics*, 29, 311-322
- A.C. Harvey, G. Phillips (1979) "Maximum likelihood estimation of regression models with ARMA disturbances," *Biometrika*, 66, 49-58
- A.C. Harvey, P.J. Todd (1983) "Forecasting economic time series with Structural and Box-Jenkins models: A case study," *Journal of Business and Economic Statistics*, 1(4)
- A.C. Harvey, A. Jaeger (1993) "Detrending, stylized facts and the business cycle," *Journal of Applied Econometrics*, 8(3), 231-247
- S.C. Hillmer, G.C. Tiao (1982) "An ARIMA model based approach to seasonal adjustment," *Journal of the American Statistical Association*, 77, 63-70
- C.C. Holt (1957) "Forecasting seasonals and trends by exponentially weighted moving averages," ONR Research Memorandum 52, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
- R.E. Kalman (1960) "A new approach to linear filtering and prediction problems". *Journal of Basic Engineering* 82 (1), 35-45
- L.R. Klein, A.S. Goldberger (1955) *An econometric model of the United States, 1929-1952*, Amsterdam
- A. Kolmogorov (1941) "Interpolation and extrapolation von stationaren Zufalligen Folgen," *Bulletin of the Academy of Sciences (Nauk), USSR, Ser. Math.*, 5, 3-14
- J. Ledolter (1984) "Comments on 'A unified view of statistical forecasting procedures' by A.C. Harvey," *Journal of Forecasting*, 3(3), 278-282
- F.R. Macaulay (1931) *The Smoothing of Time Series*, New York: NBER
- S. Makridakis, M. Hibon (1979) "Accuracy of forecasting: an empirical investigation (with discussion)," *Journal of the Royal Statistical Society A*, 142, 97-145
- I. Mintz (1969) *Dating postwar business cycles*, New York: NBER
- Muth (1960) "Optimal properties of exponential weighted moving average forecasts," *Journal of the American Statistical Association*, 55(290), pg.299
- Muth (1961) "Rational expectations and the theory of price movements," *Econometrica*, 29(3), 315-335
- C.R. Nelson, H. Kang (1981) "Spurious periodicity in inappropriately detrended time series," *Econometrica*, 49, 741-751

_____ (1984) "Pitfalls in the use of time as an explanatory variable in regression," *Journal of Business & Economic Statistics*, 2, 73-82

Nerlove and Wage (1964) "On the optimality of adaptive forecasting," *Management Science*, 10, 207-29

M. Nerlove (1967) "Distributed lags and Unobserved Components in economic time series," Ch.6 in *Ten Economic Studies in the Tradition of Irving Fisher*, W. Fellner et. al. eds., New York: John Wiley & Sons.

_____ (1972) "Lags in economic behavior," *Econometrica*, 40, 221-251

Nerlove and Grether (1970) "Some properties of "Optimal" seasonal adjustment," *Econometrica*, 38(5)

Nerlove, Grether, and Carvalho (1979) *Analysis of Economic Time Series: A Synthesis*, New York: Academic Press

P. Newbold (1983) "The competition to end all competitions," *Journal of Forecasting*, 2(3), 276-279

A. Pagan (1972) "Efficient estimation of models with composite disturbance terms," *Journal of Econometrics*, 1, 329-340

_____ (1975) "A note on the extraction of components from time series," *Econometrica*, 43, 163-168

M.B. Priestley (1981) *Spectral Analysis and Time Series*, 2 volumes, London: Elsevier Academic Press.

F. Schwappe (1965) "Evaluation of likelihood functions for Gaussian signals," *IEEE Transactions on Information Theory*, 11, 61-70

L. Taylor (1970) "The existence of optimal distributed lags," *The Review of Economic Studies*, 37, 95-106

N. Wiener (1949) *The extrapolation, interpolation and smoothing of stationary time series with engineering applications*, Wiley: New York.

P.R. Winters (1960) "Forecasting sales by exponentially weighted moving averages," *Management Science*, 6, 324-42

E. Zivot, K. Hwa Oh, D. Creal (2007) "The relationship between the Beveridge-Nelson decomposition and the Unobserved Components models with correlated shocks," Working Paper

P.C. Young (2011) *Recursive Estimation and Time-Series Analysis*, 2nd ed., Berlin: Springer-Verlag