Nonparametric Transfer Function Models with Localized Temporal Effect

Daquis, John Carlo
University of the Philippines
Diliman
Quezon City 1101, Philippines
jcdaquis@yahoo.com

Barrios, Erniel
University of the Philippines
Diliman
Quezon City 1011, Philippines
ernielb@yahoo.com

I. INTRODUCTION

Dynamism often exists on several time series wherein univariate analysis on such data is not sufficient. Box et al. (1994) discussed a function which relates the output and the input series:

$$y_t = v(B)x_t + n_t.$$

The equation above is referred to as the transfer function model. In this equation, the output series y_t is a linear function of the current and past values of the input series x_t filtered through the transfer function $v(B)X_t = \sum_{j=-\infty}^{\infty} v_j B^j$. The problem with the Box-Jenkins transfer function model is that it only accounts for linear relationships between the input and the output series. The input and output series must also be bivariate stationary. Furthermore, an inherent feature of the untransformed input and output series is they must have a certain degree of integration (Granger 1981 as cited in Granger and Engle, 2003).

A series of models were introduced to address these problems. Tjøstheim et.al. (2007), Li and Racine (2006) and Gao et.al. (2009) investigated nonlinear autoregressive transfer function models while Hidalgo (1992), Truong and Stone (1994) explored with nonparametric time series analysis.

This paper proposes a semiparametric form of the transfer function model that is flexible enough to capture both linear and nonlinear features of the input and output series. The proposed model can also accommodate nonstationarity, seasonality and structural change without going through the prewhitening methods required in the parametric transfer function modeling algorithm.

The postulated model is as follows:

$$y_t = \sum_{i=1}^{p} g_i(x_{it}) + L_t + \varepsilon_t, \;\; where\; L_t = \sum_{j=1}^{s} L_{jt} + \eta_{jt}.$$

Where:

$\{y_t\}$	=	output series
$\{x_{it}\}$	=	ith input series
$\{\varepsilon_t\}$	=	the error series
p	=	number of input variables
t	=	number of time series points
$gi(x_{it})$	=	the smooth of y_t on x_{it}
L_t	=	localized temporal effect
s	=	number of clusters
η_{jt}	=	random variation within the j^{th} cluster

II. METHODOLOGY

We decompose the output series into two components, the "baseline" output component γ_t and what is called the localized temporal component λ_t. The baseline output will be explained by the additive nonparametric function $g(x_{it})$, which will account for the dependence between the input and the output series, possibly capturing both the linear and nonlinear components.

Any perturbations in the output not found in the input (e.g. seasonality and structural changes) will be captured by the localized temporal component λ_t. In turn, this component will be modeled by the localized temporal effect L_t.

The following iterative procedure is used in estimating the model:

Step1: Smooth the output series based on the input series. Using cubic smoothing splines, estimate the smooth function $g(x_t)$ which will gives a nonparametric fit on the dependencies link between the input and the output series. The initial additive model is as follows:

$$y_t - \sum_{i=1}^{2} g(x_{it}) = \varepsilon_t$$

The innovation process ε_t is still not yet white noise since it still contains the localized temporal component.

Step 2: Compute the residuals:

$$e_t = y_t - \hat{g}(x_t)$$

This is a realization of the innovation process ε_t. The residuals ε_t contain information on the temporal dependencies left out in the output series that is not explained by the input series.

Step3: Estimate the localized temporal component with a mixed model.

Step4: Provide forecasts and compare forecast accuracy with the parametric methodology.

A simulation study is conducted to compare the postulated model with its parametric counterpart. In the simulation, a total number of 224 output time series are considered, each with 100 replicates. For the assessment of estimates, the Mean Absolute Percentage Error (MAPE) per scenario are computed for both the parametric and semiparametric procedures.

III. RESULTS AND DISCUSSION

Simulation results show that whereas the parametric procedure performs poorly for series with small time points (36), the semiparametric procedure is robust to changes in time points. As the number of points increases, The parametric procedure gives better estimates but the postulated procedure gives comparable estimates as well.

Except for the presence of ladder trend, the postulated model has consistently produced lower MAPE than the parametric procedure. Again, as the number of time points increases, the performance of the parametric procedure becomes better. For output series with structural change, the postulated model yielded lower MAPE across all time points.

Furthermore, even though the MAPE are smaller for the parametric transfer function procedure for long time series, the difference between their MAPE are not considerably large. That is, the postulated model still produces comparable estimates and is robust across the number of time points.

For small number of time points, the postulated model yielded lower MAPEs for 31 out of 56 scenarios. The postulated model has consistently performed well on estimating the second form of the transfer function model. Majority of the time series with seasonality were modeled better by the semiparametric procedure.

When there are 60 time points, the postulated model produced better estimates than the parametric procedure for majority of the scenarios with an exponentially decaying transfer function. Note also that 24 out of the 56 scenarios with 60 time points are modeled better by the semiparametric procedure.

Even though the semiparametric model produced better estimates only for 16 out of 56 scenarios for time series data with 180 data points, it has consistently produced better estimates for data with seasonality or structural change.

For large numbers of time series points, the semiparametric procedure produced lower MAPE’s only in 13 out of 56 scenarios, wherein 12 of which are for scenarios wherein seasonality, structural change or both are present.

It is also worthy to note the limitations of the parametric transfer function model. First, in all cases of localized temporal effects, the crosscorrelation function can only identify correctly the delay lag of the input series less than 50% of the time.
Except possibly for the presence of seasonality, when localized temporal effects are present in the output

series, the cross correlation function fails to specify a correct form of the transfer function. And even if a delay parameter is consistently specified, the specification is incorrect.

The postulated model, being a nonparametric one can do away with the tedious specification of the model parameters, and therefore avoiding the problem of misspecification.

Estimation convergence is also an issue for the parametric method. For data sets with a finite lag form of transfer function, 99.80% of the 11,200 estimation procedures converged. However the estimation procedure only converged for 75.75% of the 11,200 data sets with an exponentially decaying form of the transfer function.

For the semiparametric transfer function procedure, even though in invokes two highly iterative procedures, the backfitting algorithm for the estimation of the additive model and the mixed modeling procedure, none of the 22400 time series in any scenario has experienced a nonconconvergent estimation procedure.

IV. CONCLUSIONS AND RECOMMENDATIONS

The following are the conclusions in the study:

a. For time series with localized temporal effects, the parametric transfer function procedure may not be reliable.
b. The postulated model is robust with respect to the number of time series points. For short series, the semiparametric model yielded lower MAPEs than the traditional procedure. For long time series, the parametric procedure gave lower MAPEs but still the semiparametric procedure deems to be a comparable alternative.
c. The postulated model produced less MAPE than the parametric model when seasonality or structural change is present in the output time series.
d. In nonstationary input series scenarios, when a correlated input variable is added in, both procedures are affected.

There are still many things which can be added and improved in the study:

a. Consider other forms of localized temporal effects such as a time-dependent volatility structure and randomly occurring outliers.
b. Add a nonparametric autoregressive term in the postulated model.
c. Postulate methods which can detect localized temporal effects.
d. Propose a generalized nonparametric transfer function model under the foundations of the combined additive and mixed effects models.

REFERENCES

Barrios, E.B. and Vargas III, G., (2007), "Forecasting from an Additive Model in the Presence of Multicollinearity", proceedings on the 10th National Convention of Statistics, EDSA Shangri-La Hotel, October 1-2.

Box, G., Jenkins, M. and Reinsell G., (1994), Time series analysis: forecasting and control, 3rd edition, Holden-Day: San Francisco. Series G.

Buja, A., Hastie, T. and Tibshirani, R., (1989), "Linear Smoothers and Additive Models", The Annals of Statistics 17- 2, 453-510.

Campano, W. and Barrios, E.B., (2008), "Estimation on Nonstructural Time Series Model with Structural Change", a paper presented in: Joint Meeting of the 4th World Conference of the International Association for Statistical Computing (IASC) and the 6th Conference of the Asian Regional Section of the IASC, Japan.

Cook, E. and Peters, K., (1981),"The smoothing spline: a new approach to standardizing forest interior tree-ring width series for dendroclimatic studies", Tree-Ring Bulletin, 41, 45-53.

Demidenko, E., (2004), Mixed Models: Theory and Application, John Wiley and Sons Inc.:Hoboken, New Jersey.

Dominici, F., McDermott, A., Zeger, S. and Samet J., (2002), "On the Use of Generalized Additive Models in Time-Series Studies of Air Pollution and Health", American Journal of Epidemiology, 156-3, 193-203.

Gao J., (2007), "Nonlinear Time Series, Semiparametric and Nonparametric Methods", Monographs on Statistics and Applied Probability, 108, Taylor and Francis Group, LLC:USA.

Gao J., King, M., Lu, Z. and Tjøstheim, D., (2009), "Nonparametric Specification Testing for Nonlinear Time Series with Nonstationarity", Econometric Theory, 25, 1869-1892.

Granger, C. W. J. and Newbold, P. (1974), "Spurious regressions in econometrics", Journal of Econometrics, 2, 111—120.

Granger, C. and Engle, R., (2003) "Time-series Econometrics: Cointegration and Autoregressive Conditional Heteroskedasticity", advanced information on the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel, Stockholm, Sweden.

Hannan, E., (1987), "Rational Transfer Function Estimation", Statistical Science, 2-2, 135-161.

Hastie, T. and Tibshirani, R., (1990), "Generalized Additive Models", Monographs on Statistics and Applied Probability, 43, Chapman and Hall: London.

Hastie, T. and Tibshirani, R., (1995), "Generalized Additive Models", (a survey paper in Generalized Additive Models), to appear in: Encyclopedia of Statistical Sciences.

Hidalgo, F. J., (1992), "Adaptive semiparametric estimation in the presence of autocorrelation of unknown form", Journal of Time Series Analysis, 13, 47–78.

Li, Q. and Racine, J.S., (2006), Nonparametric Econometrics: Theory and Practice, Princeton University Press: Princeton.

Mammen, E., Linton, O. and Nielsen, J. (1999) "The Existence and Asymptotic Properties of a Backfitting Projection Algorithm under Weak Conditions", The Annals of Statistics, 27, 1443-1490.

Opsomer, J., (2000), "Asymptotic Properties of Backfitting Estimators", Journal of Multivariate Analysis, 73-2, 166-179.

Pierce, A., (1975), "Noninvertible Transfer Functions and their Forecasts", The Annals of Statistics, 3-6, 1354-1360.

Poskitt, D. and Tremayne, A., (1981), "An Apporach to Testing Linear time Series Models", The Annals of Statistics, 9-5, 974-986.

Riani, M., (2004), "Extensions of the Forward Search to Time Series", Linear and Nonlinear Dynamics in Time Series, 8-2, The Berkeley Electronic Press.

Shieke, T., (2001), "A Generalized Additive Model for Survival Times", The Annals of Statistics, 29-5, 1344-1360.

Tjøstheim, D., Karlsen, H. and Myklebust, T., (2007), "Nonparametric Estimation in a Nonlinear Cointegration Type Model", The Annals of Statistics, 35-1, 252-299.

Truong, Y. K. and Stone, C. J., (1994), "Semiparametric time series regression", Journal of Time Series Analysis, 15, 405–428.

Wei, W. S. (2006), Time Series Analysis: Univariate and Multivariate Models, 2nd edition, Pearson Education Inc.: USA.