On a Stepwise Hypotheses Testing Procedure and Information Criterion in Identifying Dynamic Relations between Time Series

Kasing Man[1] and Chung Chen[2]
[1]*Western Illinois University and* [2]*Syracuse University*

Abstract: This paper studies an effective stepwise hypotheses testing procedure in identifying dynamic relations between time series, and its close connection with popular information criteria such as AIC and BIC. This procedure, labeled M2, extends Chen and Lee's (1990) procedure to cover both the strong and weak form dynamic relations; and to be used with a guided choice of significance levels which are adapting in nature. Intuitively, procedure M2 can be viewed as a backward-elimination approach that simplifies the all-possible pairwise comparisons approach implied by information criterion. New insights concerning identification of strong and weak form dynamic relations using these approaches are given. Extensive simulation experiments are conducted to illustrate the performance of the IC and M2 approach in different settings. For applications, we study the dynamic relations between price level and interest rate in US and UK, and the robustness of the model identified is also addressed.

Key words: Adapting significance level, dynamic relations, information criterion, stepwise hypotheses tests, strong and weak form relations, vector ARMA model.

1. Introduction

Popular model selection criteria such as Akaike information criterion and Schwarz's Bayesian information criterion are widely used and successfully implemented in many applications. In general, using the information criterion (IC) approach, the decision for choosing an appropriate model depends upon a trade-off between the goodness-of-fit and the complexity of the model. Other commonly used criteria can be found, for example, in Billah, Hyndman and Koehler (2005). On the other hand, stepwise hypotheses testing approach is also a popular model selection strategy in specific applications. For instance, in choosing the order of vector autoregressive model, a sequential likelihood ratio tests procedure can be used to determine if the next higher order AR term is needed (*e.g.*, Reinsel

(1993, chap.4). In phylogenetics, a popular strategy to select model is hierarchical likelihood ratio tests (see, for example, Posada and Buckley (2004) and the references therein). In choosing an appropriate dynamic relation between time series, a stepwise likelihood ratio tests procedure is developed in Chen and Lee (1990).

The information criteria and hypotheses testing are apparently different approaches for model selection, and it is important to study their connection. First, the basic connection between the penalty term in information criterion and significance level in hypotheses testing is well known (*e.g.*, Smith and Spiegelhalter (1980) and Aitkin (1991)). Second, the IC approach can be viewed as a decision procedure that involves all-possible pairwise testing of the candidate models, with the critical values used in the testing approach determined by the penalty terms of the criteria (*e.g.*, Pötscher (1991)). Stoica, Selén and Li (2004) discussed an implementation of generalized likelihood ratio tests which is equivalent to IC rule and applied it for sparse models.

In this paper, we study an extended Chen and Lee's (1990) procedure and investigate its interesting and close connection to popular information criteria in the context of specifying an appropriate dynamic relation between time series.

In many business and economics applications, it is crucial to identify the presence or the absence of specific dynamic relations between the time series of interest. For instance, does full feedback relation exist between the series, or the relationship is solely uni-directional or even independent. In addition, is there concurrent association between the series or not. Following the concept of Granger's causality, such relations can be represented by imposing specific restrictions on the model parameters within the vector ARMA model. See, for example, Chen and Lee (1990) and the references therein. They proposed an effective sequential inference procedure based on a series of likelihood ratio tests on various model structures to identify the best structure. It is used to study the price and interest rate relationship and shed light on the Gibson's paradox. Chen and Wu (1999) extended the procedure to more than two series, and studied the dynamic relations among corporate dividends, earnings and prices.

A natural question arises as how does this effective procedure compare to classic model selection criterion such as AIC and BIC? We begin with extending Chen and Lee's (l990) procedure to handle dynamic relations both in the weak and strong forms in Section 2. The strong form excludes concurrent association between the series while the weak form does not. See, for example, Geweke (1986) who emphasized the importance of distinguishing the presence of concurrent relation in economics. The extended procedure, labeled M2, begins with determining an appropriate form (weak or strong) for the dynamic relations, then seeks to move up a hierarchy of models with increasingly simplified structure. A

model is selected at the simplest level the procedure can move up to, without significantly reducing the likelihood. It is important to point out that the choice of significance level to accompany the various tests is critical to the success of the procedure. To assess the effect, we perform an extensive simulation study. We experiment with a wide range of significance levels and find that a better choice in our setting is around 2.5%, which is about the usual level choice of 1 to 5%. Furthermore, when M2 is conducted at 2.5% level, the procedure gives an encouraging overall empirical probability of about 82% in correctly identifying the different types of dynamic relations under consideration. The rationale behind such preferred level choice, and a even better choice of using adapting levels, will be addressed in Section 4.

Section 3 turns to study the performance of information criterion in identifying both strong and weak form dynamic relations. To accommodate strong form relation, an effective number of parameter is used in the IC rule. Simulation results show this adjustment works well. To the best of our knowledge, this adjustment has not been studied in the literature. A variety of penalty weights including AIC, BIC, lighter and heavier weights relative to BIC are investigated in the simulation. Results in Section 3.1 show that BIC serves well in selecting the right dynamic relation overall. Its overall empirical probability in identifying the model correctly is about 84%, which is only about 2% higher than using M2. Moreover, we note that the performance of M2 when conducted at 12.5 to 15% significance level is close to AIC. This is striking as M2 and IC are apparently different approaches, but the results suggest a close connection between them in fact exists.

It is well known that when casted in the testing framework, the IC approach can be viewed as a decision procedure that involves all-possible pairwise comparisons of the candidate models (see Smith and Spiegelhalter (1980), Aitkin (1991), Pötscher (1991) and Stoica, Selén and Li (2004) for further discussions). In this light, procedure M2 when used with adapting significance levels can be viewed as a backward-elimination approach that simplifies the all-possible searching procedure, by taken into consideration model structure. A noted feature of procedure M2 is that the model under the null hypothesis is always nested within the alternative, and the number of restrictions is usually small due to the stepwise structure. This feature is appealing because the null and alternative hypotheses are not necessarily nested under the all-possible comparisons approach. See Vuong (1989) for likelihood ratio tests for non-nested hypotheses. Here, the notion that a guided choice of significance levels can be made according to the penalty term leads to further improvement of M2. Simulation studies in Section 4.1 confirm that the performance of using M2 with adapting significance levels improves to about 84%, which is essentially the same as using BIC.

For empirical applications, we revisit the price level and interest rate series in Chen and Lee (l990). We strengthen their result by further distinguishing between the strong and weak form relations in the dynamic structure. Moreover, we also illustrate how the robustness of the dynamic relation selected can be effectively assessed through the use of a variety of adapting significance levels and penalty weights.

The rest of the paper is organized as follows. Section 2 studies procedure M2 in choosing dynamic relations. Section 3 studies information criteria in choosing dynamic relations including both the strong and weak forms. Section 4 studies their connection and improves M2 by using adapting significance levels. Section 5 presents empirical applications and concludes the paper.

2. Choosing Dynamic Relations with a Stepwise Hypotheses Testing Procedure

Consider a stationary and invertible vector ARMA (VARMA) model

$$\begin{bmatrix} 1 - \phi_{11}(B) & \phi_{12}(B) \\ \phi_{21}(B) & 1 - \phi_{22}(B) \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} 1 - \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & 1 - \theta_{22}(B) \end{bmatrix} \begin{bmatrix} a_t \\ b_t \end{bmatrix}$$

where $\underline{\epsilon}_t = (a_t, b_t)'$ is an identically and independently distributed Gaussian noise with mean zero and covariance matrix Σ which consists of elements $\sigma_{i,j}, i, j = 1, 2$. The ϕ_{ij} and θ_{ij} are the usual AR and MA components and B is the back-shift operator. In many applications, it is important to determine an appropriate dynamic relation between the time series given the observed data. For instance, is the relationship uni-directional or in feedback form; and whether the errors are contemporaneously related or not.

Table 1: Parametric constraints and the corresponding hypotheses which are sufficient

Dynamic relations	Hypothesis	Parametric Constraints	
		Model parameter	Error covariance
independent (strong)	$H_\wedge : x \wedge y$	$\phi_{ij}(B) = \theta_{ij}(B) = 0$	$\sigma_{ij} = 0, i \neq j$
contemporaneous (weak)	$H_B : x \leftrightarrow y$	$\phi_{ij}(B) = \theta_{ij}(B) = 0$	
unidirectional (strong)	$H_{U_0} : x \Leftarrow_0 y$	$\phi_{21}(B) = \theta_{21}(B) = 0$	$\sigma_{ij} = 0, i \neq j$
unidirectional (weak)	$H_U : x \Leftarrow y$	$\phi_{21}(B) = \theta_{21}(B) = 0$	
unidirectional (strong)	$H_{L_0} : x \Rightarrow_0 y$	$\phi_{12}(B) = \theta_{12}(B) = 0$	$\sigma_{ij} = 0, i \neq j$
unidirectional (weak)	$H_L : x \Rightarrow y$	$\phi_{12}(B) = \theta_{12}(B) = 0$	
feedback (strong)	$H_{F_0} : x \Leftrightarrow_0 y$	no parametric constraints	$\sigma_{ij} = 0, i \neq j$
feedback (weak)	$H_F : x \Leftrightarrow y$	no parametric constraints	

These various dynamic relations under examination lead to different restrictions to be placed in the model parameters. The parametric constraints and the corresponding hypotheses which are sufficient (but not necessary) for various dynamic relations to hold are summarized in Table 1. In the table, a strong form

relation has the same structural parameterization as the weak form, and further requires the error covariance σ_{ij} to be zero. In other words, a strong form relation excludes the contemporaneous association between the series (see, for example, Geweke (1986)).

Chen and Lee (l990) developed a stepwise likelihood ratio tests procedure to identify the best dynamic specification. Their method with a minor modification, labeled as M1, is described in the Appendix using our notations. In essence, it is a backward-elimination approach that seeks to move up a hierarchy of models with increasingly simplified structures. A model is selected at the simplest level the procedure can move up to. The specifications they considered, however, do not include strong form relations. Below, we extend their procedure to cover this form.

Among the eight possible structures under the hypotheses $\{H_F, H_{F_0}, H_U, H_{U_0}, H_L, H_{L_0}, H_B, H_\wedge\}$ that we considered, it is clear that not all hypotheses between the most restrictive $H_\wedge$ and least restrictive H_F are nested. However, there are subsequences such as $\{H_F, (H_U \text{ or } H_L), H_B\}$ or its strong form counter-parts which are nested. A noted feature of our stepwise testing procedure is that the null hypothesis is always nested within the alternative.

To facilitate the procedure design, observe that if a weak form relation is estimated under the more restrictive strong form specification, the resulting log determinant of the error covariance matrix ($\ln|\Sigma|$) will be substantially affected. On the other hand, $\ln|\Sigma|$ will not be much affected if a strong form is estimated as weak form. Our limited experience from simulation studies suggested that this is the key to distinguish between the strong and weak forms. Our extended Chen and Lee's (l990) procedure begins with choosing a form (strong or weak) by testing three pairs of hypotheses (H_{L_0}, H_L), (H_{U_0}, H_U) and (H_{F_0}, H_F). If there are no significant difference between all of these strong/weak form pairs, the procedure would only search for strong form relations; otherwise it would search for weak form.

To simplify presentation, we use the notation (H_j, H_i) to denote testing of the null hypothesis H_j against the less restrictive alternative H_i using the usual likelihood ratio test. The determinant of the error covariance matrix of the VARMA model obtained under the parametric constraints imposed by hypothesis H_k is denoted by $|\Sigma_k|$. (For simplicity of notation, this is understood to be the sample estimate). The result of the likelihood ratio test is represented by an 'x' or an 'o', as judged by comparing $n*ln|\Sigma_j|/|\Sigma_i|$ to the chi-square value $\chi^2_{\#_r}(\alpha)$. The degree of freedom $\#_r$ equals to the number of parametric restrictions between the null and alternative hypotheses, α is the pre-specified significance level, and n is effective number of observations. An 'x' indicates H_j is rejected when compared to H_i at the α level; an 'o' is used if H_j is not rejected. In other words, an 'x'

indicates that $\ln|\Sigma_j|$ is significantly different from $\ln|\Sigma_i|$; an 'o' indicates that it is not. Similar notation is used for testing of more than one pair of hypotheses. For example, in testing $(H_j,\ H_i)$ and $(H_l,\ H_k)$, the result, say, (x,0), indicates that H_j is rejected in the first test and H_l is not in the second.

Procedure M2. Initial step: Test three pairs $(H_{F_0},\ H_F)$, $(H_{U_0},\ H_U)$, and $(H_{L_0},\ H_L)$. If the result is (o,o,o), go to _(Strong form) Stage one_ to test for strong form relations; otherwise, go to _(Weak form) Stage one_ to test for weak form relations.

(Strong form) Stage one: Test two pairs $(H_{U_0},\ H_{F_0})$, $(H_{L_0},\ H_{F_0})$. If the result is

1a) (x,x): conclude H_{F_0}.

1b) (x,o): conclude H_{L_0}.

1c) (o,x): conclude H_{U_0}.

1d) (o,o): go to _(Strong form) Stage two_ and check if further parametric constraints can be imposed.

(Strong form) Stage two: Test two pairs $(H_\wedge,\ H_{U_0})$, $(H_\wedge,\ H_{L_0})$. If the result is

2a) (x,o): conclude H_{U_0}.

2b) (o,x): conclude H_{L_0}.

2c) (o,o): conclude $H_\wedge$.

2d) (x,x): then test $(H_\wedge,\ H_{F_0})$. An 'x' will conclude $\min(H_{U_0},\ H_{L_0})$; an 'o', $H_\wedge$.

(Weak form) Stage one: Use Chen and Lee's (1990) procedure as described in the Appendix.

It is clear that this is not the only way to design the stepwise hypotheses testing procedure. We have also tried some alternative designs and found procedure M2 to be simple and also performed well. Moreover, the approach is intuitively appealing in the sense that the partially nested structure has been employed in the testing procedure. The other designs we tried are reported in an earlier version of the paper and are not reported here for brevity. It is important to point out that the significance level chosen to accompany the various tests in the procedure plays an important role in the success of the procedure. In the simulation study below, we experiment with a wide range of significance levels for exploration purpose. The issue of choosing a better significance level will be addressed in subsequent sections.

2.1 Simulation study: Procedure M2 conducts at various fixed significance levels

Bivariate ARMA(1,1) models are simulated according to the eight types of dynamic relations we considered. The parameter values are chosen according to the estimation results for the price level and interest rate series as studied in

Chen and Lee (l990) and their related working paper, which are given as follows:

For H_B, $\Phi = \begin{pmatrix} 1 & 0 \\ 0 & 0.92 \end{pmatrix}$, $\Theta = \begin{pmatrix} -0.3 & 0 \\ 0 & 0 \end{pmatrix}$.

For H_L, $\Phi = \begin{pmatrix} 1 & 0 \\ 0.1 & 0.92 \end{pmatrix}$, $\Theta = \begin{pmatrix} -0.3 & 0 \\ -0.3 & 0 \end{pmatrix}$.

For H_U, $\Phi = \begin{pmatrix} 1 & 0.1 \\ 0 & 0.92 \end{pmatrix}$, $\Theta = \begin{pmatrix} -0.3 & -0.3 \\ 0 & 0 \end{pmatrix}$.

For H_F, $\Phi = \begin{pmatrix} 1 & 0.1 \\ -0.2 & 0.92 \end{pmatrix}$, $\Theta = \begin{pmatrix} -0.3 & -0.3 \\ 0.3 & 0 \end{pmatrix}$.

The AR(1) and MA(1) components are $(I - \Phi B)$ and $(I - \Theta B)$ respectively. The error covariance matrix is $\Sigma = \begin{pmatrix} 9 & 3 \\ 3 & 9 \end{pmatrix}$ for weak form relations; and $\Sigma_0 = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}$ for strong form.

From each of these eight underlying models, we simulate 200 observations with the first 100 removed. This is the simulated time series, and M2 is applied to select the best specification. Specifically, we estimate eight different VARMA(1,1) models (with intercept term) to the simulated time series, each according to the restrictions imposed by the eight dynamic relations under consideration, and the corresponding sample $\ln|\Sigma|$ is recorded. The best model is then identified by M2, which is conducted at a pre-specified significance level being held fixed in all tests. In this experiment, we try a wide range of levels at 0.1, 0.5, 1, 2.5, 5, 10, 12.5 and 15% for exploration and to locate a better choice. The whole experiment is replicated 5000 times. In this experiment, a simulated series will be removed if its sample $\ln|\Sigma|$ under the H_F and $H_\wedge$ structure are not being the smallest and largest value, respectively. The number of series removed ranges from about 1% to 5% out of the total number of series simulated from each of the eight underlying models.

For each underlying model i, denote the probability of selecting the correct specification as $P_{i|i}$ (or correctly select H_i given H_i is true); and that of incorrectly selecting other specification j as $P_{j|i}$. The first panel of Table 2 reports the empirical probability of correct selection $P_{i|i}$ for each i using M2 at various significance levels. The detail tabulations of $P_{j|i}$ for each underlying model are available on request. It is useful to come up with an overall performance measure to summarize the effectiveness of M2. Here, we use a geometric mean measure defines as $\bar{P}_{all} = (\Pi_{i=1}^8 P_{i|i})^{1/8}$ to summarize the eight individual empirical probabilities of correct identification. Similarly, denote $\bar{P}_W = (\Pi_i P_{i|i})^{1/4}$, i=F, U, L, B and $\bar{P}_S = (\Pi_i P_{i|i})^{1/4}$, i=F_0, U_0, L_0, $\wedge$ as the overall performance measure for identifying the weak and strong form relations, respectively. These measures are reported in Table 3. The following observations can be made.

Table 2: Simulation studies: Empirical percentage of correctly identifying the true underlying specification using three different procedures.

(First Panel) M2 with fixed significance levels:

	Fixed level	True underlying specification							
		F	F_0	L	L_0	U	U_0	B	$\wedge$
	15%	**0.977**	0.792	0.729	0.622	0.585	0.519	0.522	0.472
	12.5%	0.973	0.824	0.756	0.668	0.617	0.566	0.567	0.522
	10%	0.964	0.856	0.774	0.717	0.652	0.622	0.614	0.580
	5%	0.928	0.918	**0.805**	0.826	0.705	0.743	0.711	0.735
M2	2.5%	0.884	0.954	0.779	0.892	**0.706**	0.802	**0.751**	0.845
	1%	0.811	**0.968**	0.692	0.915	0.636	**0.817**	0.733	0.922
	0.5%	0.740	**0.968**	0.602	**0.916**	0.562	0.802	0.684	0.955
	0.1%	0.560	0.946	0.387	0.852	0.358	0.691	0.535	**0.985**

(Second Panel) AIC and BIC with penalty adjustments v from -3.0 to+3.0:

	v	True underlying specification							
		F	F_0	L	L_0	U	U_0	B	$\wedge$
AIC		0.977	0.823	0.744	0.639	0.605	0.539	0.581	0.498
	-3.0	**0.985**	0.773	0.675	0.538	0.525	0.441	0.467	0.386
	-2.5	0.975	0.833	0.755	0.664	0.619	0.564	0.604	0.525
	-2.0	0.963	0.876	0.805	0.750	0.686	0.660	0.708	0.641
	-1.5	0.947	0.901	**0.820**	0.817	0.722	0.730	0.775	0.738
	-1.0	0.930	0.922	0.817	0.856	**0.740**	0.768	0.813	0.812
	-0.5	0.909	0.935	0.795	0.884	0.733	0.789	0.829	0.867
BIC	0	0.889	0.947	0.762	0.895	0.701	**0.794**	**0.838**	0.907
	+0.5	0.865	**0.951**	0.718	**0.896**	0.663	0.785	0.832	0.934
	+1.0	0.837	**0.951**	0.667	0.888	0.621	0.768	0.820	0.950
	+1.5	0.809	0.946	0.616	0.873	0.571	0.737	0.803	0.963
	+2.0	0.774	0.941	0.560	0.852	0.524	0.700	0.783	0.972
	+2.5	0.740	0.930	0.504	0.828	0.473	0.654	0.759	0.979
	+3.0	0.704	0.918	0.449	0.803	0.415	0.612	0.731	**0.985**

(Third Panel) M2 with adapting significance levels:

	Adapting level	True underlying specification							
		F	F_0	L	L_0	U	U_0	B	$\wedge$
	$\underline{\alpha}_A$	**0.979**	0.784	0.747	0.631	0.607	0.529	0.552	0.488
	$\underline{\alpha}_B*4$	0.971	0.819	**0.850**	0.765	0.751	0.682	0.781	0.702
	$\underline{\alpha}_B*2$	0.938	0.894	0.829	0.852	**0.762**	0.761	0.822	0.825
M2	$\underline{\alpha}_B$	0.895	0.933	0.779	0.894	0.718	**0.794**	**0.824**	0.900
	$\underline{\alpha}_B/2$	0.841	0.951	0.699	**0.901**	0.650	0.789	0.795	0.942
	$\underline{\alpha}_B/4$	0.776	**0.951**	0.609	0.887	0.566	0.759	0.745	**0.967**

Table 3: Overall measure of the empirical percentage of correctly identifying the true underlying specification using three different procedures.

(First Panel) M2 with fixed significance levels:

	Fixed level	Overall measure		
		$\bar{P}_{all}$	$\bar{P}_W$	$\bar{P}_S$
	15%	0.634	0.683	0.589
	12.5%	0.673	0.712	0.635
	10%	0.712	0.739	0.686
	5%	0.792	**0.782**	0.802
M2	2.5%	**0.823**	0.777	0.871
	1%	0.804	0.715	0.904
	0.5%	0.764	0.643	**0.908**
	0.1%	0.623	0.451	0.861

(Second Panel) AIC and BIC with penalty adjustments v from -3.0 to +3.0:

	v	Overall measure		
		$\bar{P}_{all}$	$\bar{P}_W$	$\bar{P}_S$
AIC		0.660	0.711	0.613
	-3.0	0.573	0.635	0.516
	-2.5	0.679	0.724	0.636
	-2.0	0.754	0.783	0.726
	-1.5	0.803	0.812	0.794
	-1.0	0.830	**0.822**	0.838
	-0.5	**0.840**	0.814	0.867
BIC	0	0.838	0.794	0.884
	+0.5	0.825	0.765	**0.889**
	+1.0	0.804	0.730	0.886
	+1.5	0.778	0.691	0.875
	+2.0	0.747	0.649	0.859
	+2.5	0.712	0.605	0.838
	+3.0	0.674	0.556	0.816

(Third Panel) M2 with adapting significance levels:

	Adapting level	Overall measure		
		$\bar{P}_{all}$	$\bar{P}_W$	$\bar{P}_S$
	$\underline{\alpha}_A$	0.649	0.704	0.598
	$\underline{\alpha}_B*4$	0.786	0.834	0.740
	$\underline{\alpha}_B*2$	0.833	**0.835**	0.832
M2	$\underline{\alpha}_B$	**0.839**	0.801	0.879
	$\underline{\alpha}_B/2$	0.814	0.742	**0.893**
	$\underline{\alpha}_B/4$	0.770	0.668	0.887

Results in the first panel of Tables 2 and 3 show that a better choice of significance level is 2.5%, which gives an encouraging $\bar{P}_{all} = 0.823$, or about 82% chance in selecting correctly using M2. The individual $P_{i|i}$ ranges from about 70% to 95%. When conducted at other significance levels, the performance at 1% level is fairly similar with a $\bar{P}_{all} = 0.804$, but the performance starts to decline when using levels beyond the 1 to 2.5% range.

Two interesting questions then arise naturally. First, why is 2.5% a better

choice of significance level in this setting? Second, how does procedure M2 compare to classic model selection criteria such as AIC and BIC? These issues are investigated in the next two sections.

3. Choosing Dynamic Relations with Information Criteria

By treating models with different parametric constraints as different VARMA models, the issue of identifying an appropriate dynamic relation can be addressed by model selection criteria. The IC approach provides a tool for a trade-off between the complexity of a model and the goodness-of-fit in the process of model selection. Specifically, denotes

$$IC_i = ln|\Sigma_i| + p_n \#_i / n \tag{3.1}$$

as the information criterion for model i, $\#_i$ the number of estimated parameters, and n the effective number of observations. AIC corresponds to a fixed penalty of $p_n = 2$, while BIC uses a heavier penalty term of $p_n = ln(n)$. These criteria are well studied and it is well known that BIC is consistent if the underlying true model is finite-dimensional; while AIC is motivated with a focus on prediction.

In general, many other choices of penalty term besides the popular AIC and BIC have been considered in the literature. Generally speaking, the choice of penalty term can be made according to the underlying assumptions as well as the specific objective of building model. Later on in our simulation studies, we try a wide range of p_n weights other than AIC and BIC to explore the effect of other types of information criteria. Our purpose is exploratory and we do not focus on any particular alternative choice.

Typically, the $\#_i$ in (3.1) refers to structural parameters in the model, and there is no zero restrictions placed in the error covariance matrix. Therefore, to account for strong form relations, we need to adjust the standard IC rule in (3.1). In this paper, we shall use a simple adjustment to the number of parameters in (3.1) as described below. To the best of our knowledge, this adjustment has not been studied in the literature.

Consider a k-dimensional VARMA(1,1) model where $\underline{x}_t - \Phi\underline{x}_{t-1} = \underline{\epsilon}_t - \Theta\underline{\epsilon}_{t-1}$, and $\underline{\epsilon}_t$ has unrestricted covariance matrix Σ. Decompose $\Sigma = LDL'$ where L is a lower triangular matrix with unit diagonal, and D is a diagonal matrix. Transform $\underline{x}_t$ by L^{-1}, and the original VARMA(1,1) can be re-written as $L^{-1}\underline{x}_t - L^{-1}\Phi L L^{-1}\underline{x}_{t-1} = \underline{\eta}_t - L^{-1}\Theta L\underline{\eta}_{t-1}$, where $\underline{\eta}_t = L^{-1}\underline{\epsilon}_t$ has covariance matrix D. In this form, the contemporaneous relation in the error term is now explicitly expressed as structural parameters in the model. Note that by restricting the lower diagonal elements of L^{-1} to be zero, the model becomes a VARMA(1,1) in strong form. Hence, we can regard the strong form version of a model as its weak form with restrictions, and the number of restrictions is equal to the

number of $\sigma_{ij}=0$, for $j > i$. In other words, when compared to the weak form, a strong form has $k(k-1)/2$ less parameters to estimate. In the same manner, this discussion also applies to higher order VARMA(p,q) model. Based on this rationale, we modify the definition in (3.1) to accommodate both strong and weak form relations by using the effective number of parameters as follows:

$$IC_i = ln|\Sigma_i| + p_n \#'_i/n; \quad \#'_i = \#_i - \#_\Sigma. \tag{3.2}$$

For strong form, $\#_\Sigma = k(k-1)/2$ is the number of restricted zeroes. For weak form, $\#_\Sigma = 0$ for no restrictions.

3.1 Simulation study: IC approach with a variety of penalty weights

For comparison on the same basis, we apply the IC approach to the same simulated series in Section 2.1 when we study the performance of M2. In addition to AIC $(p_n = 2)$ and BIC $(p_n = ln(n))$, we also use some lighter and heavier penalty weights for exploration purpose and to gain insight. In particular, we experiment with adjusting the BIC penalty upward and downward as $ln(n) \pm v$, where $v=0$ through 3 with an increment of 0.5. For 100 observations simulated from VARMA(1,1) model, these weights cover a wide range of penalty weights beyond AIC and BIC. The empirical probabilities $P_{i|i}$ of correct model identification using IC with this range of penalty weights and the overall measures are summarized in the second panel of Tables 2 and 3, respectively. A number of interesting observations can be made.

The simulation results show that BIC, without adjustment, serves well in selecting the correct model from the eight underlying structures as a whole. This is not surprising as our simulation is based on finite-order VARMA model and BIC is known to be consistent in selecting the right model. We also note that the strong form adjustment (3.2) works well. In addition, results show that the highest $\bar{P}_{all}$ is about 84% by using BIC or BIC with a slight downward penalty adjustment $(v = -0.5)$. Interestingly, this $\bar{P}_{all}$ is just about 2% higher than using M2 (with $\bar{P}_{all} = 0.823$). Moreover, we note that not only the performance of M2 conducts at 1 to 2.5% level is close to using BIC; M2 at 12.5 to 15% level is close to AIC. Such close connection is striking as M2 and IC are apparently different approaches. Later on in Section 4, we shall examine in details why such correspondence occurs.

While BIC performs well as a whole, further insights are observed in regards to identifying strong and weak form relations. First, it is important to point out that each dynamic structure has its own preferred penalty weight. Consider the full model H_F for instance. The second panel of Table 2 shows that the empirical probability of identifying this specification increases as v gets more negative. Intuitively, a larger downward adjustment means a lighter penalty on

the number of parameters, and therefore favours selecting the full model. In fact, our simulations show a lighter penalty weight (relative to BIC) is preferred by other weak form relations as well. Specifically, the best adjustment in identifying H_L, H_U and H_B are all non-positive, with $v_L = -1.5$, $v_U = -1$ and $v_B = 0$ respectively. In sum, for all weak form relations collectively as a group, a lighter penalty weight of $ln(n) - 1$ increases $\bar{P}_W$ to 0.822 from 0.794 of BIC's $ln(n)$ weight (see the second panel of Table 3).

On the other hand, for strong form model $H_\wedge$, a large positive adjustment is preferred. Intuitively, a high penalty $ln(n) + v$ imposes on the number of parameters favours the selection of the simplest model. In fact, the preferred adjustment for all strong form relations is non-negative, with $v_{L_0} = 0.5$, $v_{U_0} = 0$, $v_{F_0} = 0.5$ respectively. In sum, for all strong form relations collectively as a group, a heavier penalty weight of $ln(n) + 0.5$ increases $\bar{P}_S$ slightly to 0.889 from 0.884 of using BIC's $ln(n)$ weight. Even though the advantage over using BIC is minimal; further evidence can be found in the testing framework.

That each underlying specification has its own preferred penalty adjustment is also noted in the testing framework when M2 is used, but is now manifested as a preferred choice of significance level. Results in the first panel of Table 3 shows that a larger significance level is preferred in identifying weak form relations ($\bar{P}_W$=0.782 using 5% improves slightly over $\bar{P}_W$=0.777 of using 2.5%); while a smaller level is preferred in identifying strong form relations ($\bar{P}_S$=0.908 of using 0.5% improves over $\bar{P}_S$=0.871 of using 2.5%). Intuitively, a larger/smaller significance level makes the procedure easier/harder to reject the more restrictive null hypothesis and conclude the less restrictive alternative. This is consistent with imposing a lighter/heavier penalty on the number of parameters. Further discussion will be given in Section 4.

To summarize, our simulation results support BIC as a good overall criterion in identifying the eight dynamic relations of interest and no major advantage is seen by using a lighter or heavier penalty. However, if prior information about the form (strong or weak) of the dynamic relation is available, then penalty/significance level can be adjusted accordingly as described above to favour the identification of that type of relations, and this is achieved at the expense of the other type.

This section illustrates how the penalty weight p_n and the number of parameters $\#_i$ in the standard information criterion for VARMA model are subjected to possible adjustment. This occurs when prior information about the model type is available and when there is zero restrictions in Σ, respectively. Formal investigation of these adjustments is interesting future work.

4. The Connection between Information Criterion and Procedure M2

The connection between the penalty term p_n in information criterion and significance level α in likelihood ratio test is well known (*e.g.*, Smith and Spiegelhalter (1980) and Aitkin (1991)):

$$p_n \#_r = \chi^2_{\#_r}(\alpha).$$

For convenience, its rationale is outlined as follow. Consider two VARMA models that model 1 is nested within model 2. Using the minimum IC rule, if $IC_2 - IC_1 = ln|\Sigma_2|/|\Sigma_1| + p_n \#_r/n > 0$ or $n * ln|\Sigma_1|/|\Sigma_2| < p_n \#_r$, model 1 will be selected. On the other hand, we can use likelihood ratio test to test model 1 as the null hypothesis against model 2 as the alternative. Model 1 is concluded if $n * ln|\Sigma_1|/|\Sigma_2| < \chi^2_{\#_r}(\alpha)$, where $\chi^2_{\#_r}(\alpha)$ is the chi-square value with degree of freedom $\#_r = \#_2 - \#_1$ at the α level. Therefore, if α is chosen as a function of $\#_r$ and p_n, such that it solves $p_n \#_r = \chi^2_{\#_r}(\alpha)$, the likelihood ratio test would conclude the same as using minimum IC.

This connection implies that the minimum IC approach can be viewed as a decision procedure that involves pairwise comparisons of all candidate models under examination. See Pötscher (1991) and Stoica, Selén and Li (2004) for details and how the critical values in the testing framework are determined by the penalty terms of the criteria. For convenience, the key characteristic of this IC-induced all-possible pairwise testing procedure is summarized as follows. The idea is to find a specification H_* which is 'close' to all other specifications H_i with more parameters (i.e., $n * ln|\Sigma_*|/|\Sigma_i| < p_n(\#_i - \#_*) = \chi^2_{\#_r}(\alpha)$); yet 'far away' from all other specifications H_j with less parameters (i.e., $n * ln|\Sigma_j|/|\Sigma_*| > p_n(\#_* - \#_j) = \chi^2_{\#_r}(\alpha)$); and is the 'best' (i.e., $n * ln|\Sigma_k|/|\Sigma_*| > 0$) within its own class ($\#_k = \#_*$). The 'closeness' (or significance) is being judged by the critical value $p_n \#_r$, or its implied significance level $\alpha(\#_r, p_n)$ which is not fixed. In this testing framework, such adapting significance level $\alpha(\#_r, p_n)$ plays the same critical role as the penalty term p_n in information criterion. Some of the likelihood ratio tests involved in the IC-induced testing procedure are not standard because the null and alternative hypotheses are not necessarily nested.

In accordance with the above interpretation, procedure M2 can be viewed as a backward-elimination procedure that simplifies the all-possible comparisons procedure implied by minimum IC approach. In other words, procedure M2 captures the essence of the IC-induced procedure by seeking the 'closest' simplified model (without significantly reducing the likelihood) in a sequential manner, starting from the full feedback model. Using the same idea, one can also design a forward-addition approach similar to M2 but begins the search with the simplest $H_\wedge$ model. The performance of M2 when used with adapting significance levels

is expected to be similar to the IC approach. This is illustrated by simulations next.

4.1 Simulation study: Procedure M2 with adapting significance levels

We apply the testing procedure M2 with adapting significance level to the simulated series in Section 2.1 and 3.1, and results are compared to that of using fixed levels as well as the IC approach.

In our simulation set-up, it is easy to check that the adapting significance levels corresponds to AIC is $\underline{\alpha}_A$=(0.1573, 0.1353, 0.1116, 0.0916) for 1 to 4 restrictions respectively; and the adapting levels corresponds to BIC is $\underline{\alpha}_B$=(0.0321, 0.0101, 0.0032, 0.0010) when the effective sample size is 99. Note that for tests involve one or two parameter differences, BIC corresponds to a significance level of about 3% and 1% and AIC corresponds to about 16% and 14%. Recall that many of the hypotheses tests in M2 involve only 1 or 2 parameter differences. Hence, if a fixed significance level were to use in all tests, a level of 1 to 2.5% will generally perform like BIC; and a 12.5% to 15% level perform like AIC. This provides a justification to the findings we had in Section 2.1 and 3.1. Obviously, the choice of a better fixed significance level will depend on the sample size as well as the most common $\#_r$ used and therefore may change in different applications.

Just as we use a variety of penalty weights p_n for exploration in Section 3.1, we also try a variety of adapting significance levels here. We explore with two proportionally larger levels (2 and 4 times of $\underline{\alpha}_B$), as well as two proportionally smaller levels (1/2 and 1/4 times of $\underline{\alpha}_B$) to gain insight and assess the sensitivity of the model selected. The chance of correct model selection under each specification $P_{i|i}$ and the overall measure are summarized in the third panel of Tables 2 and 3, respectively. In comparison to the results of using IC approach, it clearly shows that M2 with adapting significance level $\underline{\alpha}_B$ performs similarly to BIC; and $\underline{\alpha}_A$ to AIC. Their chances of selecting over the eight underlying models follow very similar pattern. The highest $\bar{P}_{all} = 0.839$ of using $\underline{\alpha}_B$ is essentially the same as 0.838 of using BIC.

For individual specification, the remark we made in Section 3.1 that each specification has its own preferred penalty adjustment is again noted here. In identifying the full feedback model H_F, AIC or $\underline{\alpha}_A$ is preferred. AIC imposes a lighter penalty on the number of parameters, and therefore tends to favour model with more parameters. Similarly, using a larger significance level $\underline{\alpha}_A$ (relative to $\underline{\alpha}_B$) is easier to reject the null hypothesis and conclude the less restrictive alternative. Intuitively, this means the procedure is harder to move up the hierarchy of models with simplified structures, thus favouring the selection of the full model. On the contrary, a heavier penalty term or a smaller significance level favours the identification of the simplest model $H_\wedge$.

In between these two extremes $H_\wedge$ and H_F, broadly speaking, weak form relations generally prefer a relatively large significance level or light penalty term; and the reverse is true for strong form relations. Consider H_L for example. The use of $\underline{\alpha}_B$ or BIC correctly identifies the true model about 78% of the time, but increases to about 85% when 4 times $\underline{\alpha}_B$ were used. For weak form relations collectively as a group, a relatively large significance level is preferable ($\bar{P}_W$=0.835 of using 2 times $\underline{\alpha}_B$ improves over $\bar{P}_W$=0.801 of using $\underline{\alpha}_B$); while the reverse is true for strong form relations ($\bar{P}_S$=0.893 of using 0.5 times $\underline{\alpha}_B$ improves over $\bar{P}_S$=0.879 of using $\underline{\alpha}_B$). Recall that procedure M2 begins with selecting a form by testing three strong and weak form pairs. Hence, if the true structure is in strong (weak) form, it is desirable to use a smaller (larger) significance level or a heavier (lighter) penalty weight as found in Section 3.1 because it would make the strong form null hypothesis harder (easier) to reject.

5. Empirical Analysis: Prices and Interest Rates

For empirical analysis, we revisit the price (P_t) and interest rate (I_t) dynamics studied in Chen and Lee (l990). They examined the full sample period as well as two sub-periods selected according to monetary standard. Using the various approaches described in this paper, Tables 4 and 5 report the model selection results for the US and UK series respectively. The conclusions we draw confirm their results and that all $P \Rightarrow I$ relations found in US second and full periods as well as UK full period can be strengthened to $P \Rightarrow_0 I$. Further details regarding the performance of various procedures as well as the robustness of the structure selected are discussed below.

In Tables 4 and 5, Column 2 reports the $\ln|\Sigma_i|$ when the model is estimated under the parametric restriction imposed by H_i. To simplify the table, we do not list all likelihood ratio comparisons made in our procedure, but only present those with respect to the full model H_F. In particular, $n * ln(|\Sigma_i|/|\Sigma_F|)$ of a hypothesis H_i from the base H_F is reported in column 3. The model selected by M2 using level $\underline{\alpha}_A$ is marked by a '‡'. For the adapting level $\underline{\alpha}_B$ as well as two proportionally higher and lower levels (4, 2, 1, 0.5, 0.25 times of $\underline{\alpha}_B$), the model selected by k of these levels are marked by a '$\star_k$'. Details regarding the corresponding k levels is described in the text. Columns 4 and 5 report the AIC and BIC values, and the model selected accordingly is marked by a '*'. We also try different penalty adjustments to the BIC weight as $p_n(v) = ln(n) + v$, where v ranges from -3 to 3 at an increment of 0.5. For those v that select differently from BIC, the model they selected is marked by an '∘', and details regarding the corresponding v is described in the text.

Table 4: US price (P_t) and interest Rate (I_t): Model selection using M2 and IC approach

Model	$\ln|\Sigma_i|$	$n^*ln(\frac{|\Sigma_i|}{|\Sigma_F|})$	AIC	BIC
US: Full period from 1:182, VARMA(1,1) fit				
H_F	-10.062	0.000	-9.952	-9.775
H_{F_0}	-10.061	0.181‡	-9.962*	-9.803
H_L	-10.039	4.163	-9.951	-9.809
H_{L_0}	-10.038	**4.344**$\star_5$	-9.961	**-9.837***
H_U	-9.940	22.082	-9.852	-9.710
H_{U_0}	-9.940	22.082	-9.863	-9.739
H_B	-9.924	24.978	-9.858	-9.752
$H_\wedge$	-9.924	24.978	-9.869	-9.780
Conclusion:	$P \Rightarrow_0 I$			
US: Period 2 from 115:182, VARMA(1,1) fit				
H_F	-9.777	0.000	-9.478	-9.149
H_{F_0}	-9.775	0.134	-9.506	-9.210∘
H_L	-9.739	2.546	-9.500	-9.237
H_{L_0}	-9.738	**2.613**‡,$\star_5$	-9.529*	**-9.299***
H_U	-9.465	20.904	-9.226	-8.963
H_{U_0}	-9.464	20.971	-9.255	-9.025
H_B	-9.432	23.115	-9.253	-9.055
$H_\wedge$	-9.432	23.115	-9.283	-9.118
Conclusion:	$P \Rightarrow_0 I$			
US: Period 1 from 75:114, VARMA(1,0) fit				
H_F	-11.349	0.000	-11.041	-10.785
H_{F_0}	-11.325	0.936	-11.069	-10.855
H_L	-11.345	0.156	-11.089	-10.875∘
H_{L_0}	-11.321	1.092‡,$\star_2$	-11.116*	-10.945∘
H_U	-11.258	3.549	-11.002	-10.788
H_{U_0}	-11.234	4.485	-11.029	-10.858
H_B	-11.250	3.861	-11.045	-10.874
$H_\wedge$	-11.229	**4.680**$\star_3$	-11.075	**-10.947***
Conclusion:	$P \wedge I$			

NOTE. For US series, full period is 1800-1981, sub-period 2 is 1914-1981, and sub-period 1 is 1874-1913. In column 3, the notations ‡ and $\star_k$ denote the models chosen by M2 at significance levels $\underline{\alpha}_A$, $\underline{\alpha}_B$, and some proportionally larger and smaller levels of $\underline{\alpha}_B$. The notation * in column 4-5 denotes the models chosen by AIC and BIC; and notation ∘ denotes the model selected by those adjustments v that is not the same as BIC (the specific v are given in the text). For full period and sub-period 2, all variety of $\underline{\alpha}_B$ chose the same model. For sub-period 1, all variety of $\underline{\alpha}_B$ chose $H_\wedge$ except for 4 and 2 times of $\underline{\alpha}_B$, which chose H_{L_0}.

Table 5: UK price (P_t) and interest rate (I_t): Model selection using M2 and IC Approach

Model	$\ln|\Sigma_i|$	$n^* ln(\frac{|\Sigma_i|}{|\Sigma_F|})$	AIC	BIC
UK: Full period from 1:253, VARMA(1,1) fit				
H_F	-10.469	0.000	-10.390	-10.250
H_{F_0}	-10.455	3.528	-10.384	-10.258
H_L	-10.465	1.008‡,$\star_1$	-10.402*	-10.289∘
H_{L_0}	-10.451	**4.536$\star_4$**	-10.395	**-10.297***
H_U	-10.382	21.924	-10.319	-10.206
H_{U_0}	-10.368	25.452	-10.312	-10.214
H_B	-10.375	23.688	-10.327	-10.243
$H_\wedge$	-10.364	26.460	-10.324	-10.254
Conclusion:	$P \Rightarrow_0 I$			
UK: Period 2 from 186:253, VARMA(1,1) fit				
H_F	-9.577	0.000	-9.278	-8.949
H_{F_0}	-9.533	2.948	-9.264	-8.968
H_L	-9.556	1.407‡,$\star_1$	-9.317*	-9.054∘
H_{L_0}	-9.511	4.422	-9.302	-9.072
H_U	-9.484	6.231	-9.245	-8.982
H_{U_0}	-9.440	9.179	-9.231	-9.001
H_B	-9.458	7.973	-9.279	-9.081
$H_\wedge$	-9.417	**10.720$\star_4$**	-9.268	**-9.103***
Conclusion:	$P \wedge I$			
UK: Period 1 from 93:185, VARMA(2,0) fit				
H_F	-12.620	**0.000‡,$\star_3$**	-12.400*	**-12.124***
H_{F_0}	-12.483	12.467	-12.285	-12.037
H_L	-12.514	9.646$\star_2$	-12.338	-12.117∘
H_{L_0}	-12.377	22.113	-12.223	-12.030
H_U	-12.468	13.832	-12.292	-12.071
H_{U_0}	-12.331	26.299	-12.177	-11.984
H_B	-12.337	25.753	-12.205	-12.040
$H_\wedge$	-12.225	35.945	-12.115	-11.977
Conclusion:	$P \Leftrightarrow I$			

NOTE. For UK series, full period is 1729-1981, and sub-period 2 is 1914-1981, sub-period 1 is 1821-1913. In column 3, the notations ‡ and $\star_k$ denote the models chosen by M2 at significance levels $\underline{\alpha}_A$, $\underline{\alpha}_B$, and some proportionally larger and smaller levels of $\underline{\alpha}_B$. The notation * in columns 4-5 denotes the models chosen by AIC and BIC; and the notation ∘ denotes the model selected by those adjustments v that is not the same as BIC (the specific v is given in the text). For full period and sub-period 2, all variety of $\underline{\alpha}_B$ chose the same model except for 4 times of $\underline{\alpha}_B$, which chose H_L. For sub-period 1, all variety of $\underline{\alpha}_B$ chose H_F except for 1/4 and 1/2 times of $\underline{\alpha}_B$, which chose H_L.

As expected, procedure M2 with levels $\underline{\alpha}_A$ and $\underline{\alpha}_B$ selected similarly as AIC and BIC respectively. For the US series, strong form dynamic relationship is supported over the full and both sub-periods. In sub-period 2, we note that the H_{L_0} structure identified is robust. This is the model chosen by M2 with all 6 different adapting significance levels, as well as AIC, BIC and BIC with all adjustments (except for $v = -3$ which chooses H_{F_0}, but a large negative adjustment is not preferred for strong form relation). Also, the likelihood ratio

test statistics show an obvious jump after model H_{L_0} as the model structure attempts to simplify further.

In sub-period 1, we also select $H_\wedge$ as Chen and Lee (l990). First note that the largest likelihood ratio test statistic among all models is just 4.68 for $H_\wedge$ which is not 'far' from the full feedback model. Also, model $H_\wedge$ is chosen by BIC and M2 conducts at $\underline{\alpha}_B$ and all relatively small levels, and they are the preferred levels to identify strong form relations. Moreover, the model selected is robust using BIC and all its positive adjustments (all $v \geq 0$). BIC with negative adjustments $(-2.5 \leq v \leq -0.5)$ choose H_{L_0}, and when a even larger negative adjustment of $v = -3$ is used, H_L is chosen. However, a negative adjustment or small penalty weight is not preferred in identifying strong form relations. Our intuition is that the use of a smaller penalty or larger significance level makes the procedure harder to move up the hierarchy of models. Here, the procedure basically stops at H_{L_0} before reaching $H_\wedge$. In this sense, we regard H_{L_0} as a second choice to $H_\wedge$. This is noteworthy as H_{L_0} is the structure selected in sub-period 2.

For the full period, we choose H_{L_0} over H_{F_0} since the use of AIC or $\underline{\alpha}_A$ is not preferred in identifying H_{F_0}. Also, BIC with all adjustments $(-3 \leq v \leq 3)$ choose the same model H_{L_0}. The choice of H_{L_0} structure for the full period is reasonable, as judged by the structures identified over sub-period 1 and 2, which are $H_\wedge$ (with H_{L_0} as a second choice) and H_{L_0} respectively. To summarize, we find that US price and interest rate do not have concurrent association (or in strong form) over the full period as well as the two sub-periods chosen according to monetary standard. Interest rate is related linearly to the previous prices but not the other way around, except for sub-period 1 when the relationship is simply independent.

Turning to the UK prices and interest rates, we also find (see Chen and Lee (l990)) that the strong form relation is appropriate to sub-period 2 but not for sub-period 1. For sub-period 1, we choose the full feedback model H_F because it is selected by AIC, BIC (and with all negative v), and M2 with larger significance levels; these are the preferred criteria to identify weak form relation. The BIC with positive adjustments $(v \geq 0.5)$ and M2 with smaller level choose H_L. Our intuition is that under such criteria, the procedure is easier to move up the hierarchy of models and in this case reaching H_L from H_F. For sub-period 2, the independent structure $H_\wedge$ is chosen, with H_L as a second choice. The H_L is chosen when a larger significance level or lighter penalty $(v \leq -1.5)$ relative to BIC is used. Again, these criteria make the procedure harder to move up, and in this case it stops at H_L before reaching $H_\wedge$. For the full period, the results favour H_{L_0} using BIC and M2 with various significance levels except $4\underline{\alpha}_B$. A second choice would be H_L, which is chosen by using the larger level and BIC with large negative adjustments $(v \leq -2.5)$.

To summarize, we confirm that contrary to US, price and interest rate in UK are independent in sub-period 2, whereas a full feedback relation is detected in sub-period 1. The selection of a strong from H_{L_0} structure in the full period while having a weak form H_L structure as second choice reflects the difficulty of compromising between the vastly different behaviours over the two sub-periods. In fact, the preference of having strong form relation over weak form is not overwhelming. Finally, by exploring with different adapting significance levels and penalty adjustments, we find that H_L is a common second choice for both sub-periods and full period. This consistency is notable in view of the vastly different dynamic relations selected over different sub-periods.

As a conclusion, this paper sheds new lights on a stepwise hypotheses testing procedure and information criteria in choosing an appropriate dynamic relation between time series. The simulation studies and empirical applications illustrate the similarity between the IC approach and M2 conducts at adapting significance levels. Moreover, we illustrate how the use of BIC with penalty adjustments or M2 with a variety of adapting levels helps assess the robustness and sensitivity of the model selected, and develop further insights regarding the underlying structures. The analysis performed in this paper will be useful in studying other similar stepwise hypotheses testing procedures.

Acknowledgements

Appendix. Procedure M1

Using our notations, the following procedure, labeled as M1, describes Chen and Lee's (l990) method with a minor modification.

Procedure M1, Stage one. Test two pairs (H_U, H_F), (H_L, H_F). If the result is

1a) (x,x): conclude H_F.
1b) (x,o): conclude H_L.
1c) (o,x): conclude H_U.
1d) (o,o): go to *stage two* and check if further parametric constraints can be imposed.

Procedure M1, Stage two. Test two pairs (H_B, H_U), (H_B, H_L). If the result is

2a) (x,o): conclude H_U.
2b) (o,x): conclude H_L.

2c) (o,o): then test $(H_\wedge, H_B)$. An 'x' will conclude H_B; an 'o', $H_\wedge$.

2d) (x,x): then test (H_B, H_F). An 'x' will conclude $\min(H_U, H_L)$ (the original method conclude H_F). If the result is an 'o', then test $(H_\wedge, H_B)$. An 'x' will conclude H_B; an 'o', $H_\wedge$.

References

Aitkin, M. (1991). Posterior bayes factors. *Journal of the Royal Statistical Society* B **53**, 111-142.

Billah, M. B., Hyndman, R. J. and Koehler, A. B. (2005). Empirical information criteria for time series forecasting model selection. *Journal of Statistical Computation and Simulation* **75**, 831-840.

Chen, C. and Lee, C. (1990). A VARMA test on the Gibson paradox. *The Review of Economics and Statistics*, Vol. LXXII **1**, 96-107.

Chen, C. and Wu, C. (1999). The dynamics of dividends, earnings and prices: evidence and implications for dividend smoothing and signaling. *Journal of Empirical Finance* **6**, 29-58.

Geweke, J. (1986). The superneutrality of money in the United States: an interpretation of the evidence. *Econometrica* **54**, 1-21.

Posada, D. and Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**, 793-808.

Pötscher, B. M. (1991). Effects of model selection on inference. *Econometric Theory* **7**, 163-185.

Reinsel, G.C. (1993). *Elements of Multivariate Time Series Analysis*. Springer-Verlag.

Smith, A.F.M. and Spiegelhalter, D.J. (1980). Bayes factors and choice criteria for linear models, *Journal of the Royal Statistical Society* B **42**, 213-220.

Stoica, P. Selén, Y. and Li, J. (2004). On information criteria and the generalized likelihood ratio test of model order selection. *IEEE Signal Processing Letters* **11**, 794-797.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307-333.

Kasing Man
College of Business and Technology
Western Illinois University
Macomb, IL 61455
KS-Man@wiu.edu

Chung Chen
Whitman School of Management
Syracuse University
Syracuse, NY 13244
cchung@syr.edu