Université de Montréal

Inférence exacte simulée et techniques d'estimation dans les modèles VAR et VARMA avec applications macroéconomiques

par

Tarek Jouini

Département de sciences économiques

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de

Philosophiae Doctor (Ph.D.)

en sciences économiques

Janvier 2008

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée :

Inférence exacte simulée et techniques d'estimation dans les modèles VAR et VARMA avec applications macroéconomiques

présentée par

Tarek Jouini

a été évaluée par un jury composé des personnes suivantes :

Benoit Perron

président-rapporteur

Jean-Marie Dufour

directeur de recherche

Marine Carrasco

membre du jury

John W. Galbraith, Univ. McGill

examinateur externe

Christian Léger

représentant du doyen de la FES

Sommaire

Cette thèse traite plusieurs thèmes liés à l'estimation et à l'inférence économétrique dans les modèles autorégressifs multivariés (VAR) ainsi que les modèles autorégressifs-moyenne-mobile multivariés (VARMA), en insistant sur les applications macroéconomiques. L'inférence statistique dans de tels modèles est typiquement basée sur des approximations asymptotiques qui peuvent être très peu fiables dans les échantillons finis ou encore sont difficiles à estimer à cause de problèmes associés à la non linéarité (modèles VARMA). Cette thèse cherche à faire progresser la solution de ces problèmes dans deux voies principales. Premièrement (dans le premier essai), nous montrons comment une inférence à distance finie peut être implémentée dans les modèles VAR sous des hypothèses paramétriques, au moyen de techniques des tests de Monte Carlo Maximisés (MCM) [Dufour (2006)]. Deuxièmement (dans les essais 2 et 3), nous développons des techniques d'estimation relativement simples pour les modèles VARMA sous la représentation forme échelon (qui permet une identification unique du modèle), qui pourront être facilement exploitées pour fin d'inférence dans ces modèles, dans le cadre des techniques du bootstrap et des tests MCM.

Dans le premier essai, nous soulignons le fait que l'inférence statistique dans les modèles VAR est typiquement basée sur des approximations, de grands échantillons, basées sur des lois asymptotiques ou des techniques du bootstrap. Après avoir démontré que de telles méthodes peuvent être très peu fiables pour des tailles d'échantillon réalistes, spécialement lorsque le nombre de retards ou le nombre d'équations est grand, nous proposons une technique générale basée sur la simulation qui permet de contrôler complètement le niveau des tests dans les modèles VAR paramétriques. En particulier, nous montrons que les tests MMC permettent de construire des tests exacts dans de tels modèles (stationnaires ou intégrés). Des tests sur l'ordre du VAR ainsi que des tests d'hypothèses de non causalité sont considérés comme cas spéciaux. La technique développée est appliquée à des modèles VAR de l'économie américaine.

Dans les second et troisième essais, nous reconsidérons l'estimation des modèles VARMA. Cette classe de modèles qui comprend les modèles VAR comme cas spéciaux,

peut fournir des représentations plus parcimonieuses de la structure dynamique de séries temporelles multivariées, et de là des estimations plus précises des paramètres d'intérêt (i.e., les coefficients d'impulsion) et de meilleures prévisions. Cependant, les modèles VARMA sont non linéaires et peuvent être très difficiles à estimer. Dans le deuxième essai, nous étudions la distribution asymptotique d'un estimateur linéaire simple de deux étapes pour des modèles VARMA stationnaires et inversibles sous la forme échelon, avec des ordres connus. Des conditions générales pour la convergence ainsi que la normalité asymptotique sont fournis. Un estimateur convergent de la matrice de variance-covariance asymptotique de l'estimateur est également proposé, rendant ainsi aisée, la construction des tests et intervalles de confiance. Enfin, dans le troisième essai, nous proposons des estimateurs linéaires fortement convergents pour les modèles VARMA sous la forme échelon, avec des ordres inconnus. La forme échelon requiert la spécification des indices de Kronecker qui sont typiquement estimés en minimisant un critère d'information de type Akaike, tel que les critères proposés par Hannan et Rissanen (1982), Hannan et Kavalieris (1984b) et Poskitt (1992). Ainsi la première étape de notre procédure consiste à proposer plusieurs améliorations de ces derniers critères – incluant des facteurs de pénalité qui ne tendent pas rapidement vers zéro lorsque la taille de l'échantillon tend vers l'infini – ce qui permet d'obtenir des estimateurs convergents des indices de Kronecker et d'augmenter la probabilité d'estimer les vraies valeurs de ces indices dans les petits échantillons. Dans une deuxième étape, on peut alors obtenir des estimateurs linéaires convergents (inefficaces). Enfin, dans une troisième étape, on peut calculer, par des méthodes linéaires, un estimateur qui est asymptotiquement équivalent au maximum de vraisemblance, et par conséquent asymptotiquement efficace. La performance de la méthode proposée est étudiée et démontrée dans une étude de simulation. Finalement, la procédure est appliquée à un modèle VARMA de l'économie américaine.

Mots clés : VAR ; tests exacts ; tests MCM ; causalité au sens de Granger ; séries temporelles ; modèles intégrés ; VARMA ; stationnaire ; inversible ; forme échelon ; indices de Kronecker ; Hannan-Rissanen ; estimation ; Monte Carlo ; simulation ; bootstrap.

Summary

This dissertation studies a number of topics related to estimation and inference on vector autoregressive (VAR) models and vector autoregressive moving average (VARMA) models in econometrics, with an emphasis on macroeconomic applications. Inference procedures for such models are typically based on large-sample approximations, which can be very unreliable in finite samples or difficult to implement because of nonlinearities (VARMA models). The dissertation makes progress on these difficulties along two broad avenues. First (in essay 1), we show how finite-sample inference can be achieved in VAR and VARMA models under parametric assumptions, through the use of maximized Monte Carlo (MMC) test techniques [Dufour (2006)]. Second (in essays 2 and 3), we develop relatively simple estimation methods for VARMA models in the echelon form representation (used to uniquely identifying the model), which might easily be exploited for inference purpose to obtain computationally inexpensive bootstrap and Monte Carlo test methods for such models.

In the first essay, we stress that statistical inference in VAR models is typically based on large-sample approximations, involving the use of asymptotic distributions or bootstrap techniques. After documenting that such methods can be very misleading even with realistic sample sizes, especially when the number of lags or the number of equations is not small, we propose a general simulation-based technique that allows one to control completely the level of tests in parametric VAR models. In particular, we show that MMC tests can provide provably exact tests for such models, whether they are stationary or integrated. Tests on the order of the VAR, and of non-causality hypotheses are considered as special cases. The technique developed is applied to a VAR model of the U.S. economy. This paper has been published in *Journal of Econometrics* [see Dufour and Jouini (2006)].

In the second and third essays, we reconsider the estimation of VARMA models. This class of models which includes as a special case the widely used VAR models, can lead to considerably more parsimonious representations of the dynamic structure of a multivariate time series, hence to more precise estimates of parameters of interest (e.g.,

impulse responses) and forecasts. On the other hand, VARMA models are nonlinear and can be quite difficult to estimate. In the second essay, we study the asymptotic distribution of a simple two-stage linear estimator for stationary invertible VARMA models in echelon form, with known orders. General conditions for consistency and asymptotic normality are given. A consistent estimator of the asymptotic covariance matrix of the estimator is also provided, so that tests and confidence intervals can easily be constructed. This paper also has been published as a chapter in a book on statistical modeling and analysis of complex data problems, edited by *Springer-Verlag* [see Dufour and Jouini (2005)]. Finally, in the third essay, we propose linear strongly consistent estimators of stationary invertible VARMA models in echelon form of unknown order. The echelon form is used to uniquely identify the parameters of the model and requires specifying order parameters called Kronecker indices. The latter are usually estimated by minimizing an information (Akaike-type) criterion, such as the criteria proposed by Hannan and Rissanen (1982), Hannan and Kavalieris (1984b) and Poskitt (1992). So the first step of our procedure consists in proposing various improvements to the latter criteria – involving penalty corrections that do not quickly vanish as the sample size goes to infinity – which provide consistent estimates of the Kronecker indices and increase the probability of estimating the true Kronecker indices in finite samples. In a second step, consistent (inefficient) linear estimators may then be obtained. A third step, requiring again only linear estimation, yields estimators which are asymptotically equivalent to maximum likelihood estimators, hence asymptotically efficient. The performance of the proposed method is studied and demonstrated in a simulation study. Finally, the procedure is applied to a small VARMA model of the U.S. economy.

Key words : VAR; exact tests; MMC tests; Granger causality; time series; integrated models; VARMA; stationary; invertible; echelon form; Kronecker indices; Hannan-Rissanen; estimation; Monte Carlo; simulation; bootstrap.

Table des matières

Liste des hypothèses, propositions et théorèmes

Liste des tableaux

Table des figures

Dédicace

À Mohammed-Rayen et Sarra

Remerciements

Cette thèse n'aurait pu aboutir sans la patience et la foi. Je tiens en premier lieu à remercier mon directeur de recherche Jean-Marie Dufour pour son aide, ensuite les organismes de recherche académiques sans lesquels le financement de ce projet de recherche n'aurait pu être assuré, à savoir le centre de recherche et développement en économique (C.R.D.E.) et le centre interuniversitaire de recherche en analyse des organisations (C.I.R.A.N.O.).

Enfin, j'adresse un grand merci à ma femme et aux membres de ma famille pour le soutien moral qu'ils m'ont offert avec générosité tout au long de ce travail.

Introduction générale

L'inférence économétrique dans les modèles à séries chronologiques multivariées souffre d'un problème de fiabilité qui a trait aux techniques d'estimations et/ou d'inférence utilisées dans ces modèles, dont on peut citer à titre d'exemples les modèles VAR (autorégressifs multivariés) et les modèles VARMA (autorégressifs et à moyenne mobile, multivariés). Dans ces modèles, souvent désignés par dynamiques, ce problème s'accentue davantage avec les échantillons de petites tailles.

En particulier, les modèles VAR sont largement utilisés dans l'analyse économétrique de la structure dynamique des séries chronologiques; voir Sims (1980), Lütkepohl (1993, 2001), Reinsel (1993), Hamilton (1994), Hendry (1995), Gouriéroux et Monfort (1997), Dhrymes (1998) et Clements et Hendry (2002). Une des raisons de cette popularité est que ces modèles sont faciles à estimer et peuvent rendre compte de la complexité dynamique des données. D'importantes applications utilisant les modèles VAR, portent sur l'analyse des prévisions [les fonctions d'impulsion ou de réaction], de causalité [au sens de Wiener (1956) et Granger (1969)], de cointégration, etc.

Cependant, les modèles VAR impliquent souvent un grand nombre de paramètres. Ce qui complique davantage les difficultés associées à l'inférence statistique dans ces modèles, surtout avec des dimensions élevées. De plus, il est largement reconnu que l'inférence statistique (tests et intervalles de confiance) dans ces modèles est typiquement basée sur des approximations distributionnelles asymptotiques. En l'occurrence, même dans les modèles statiques de régression linéaire multivariés, il a été reconnu que les approximations asymptotiques peuvent être très peu fiables; voir Dufour et Khalaf (2002). Ces problèmes s'aggravent davantage dans les modèles VAR (qui peuvent être interprétés aussi comme étant des modèles de régression linéaire multivariés dynamiques), même sous des hypothèses fortes de régularité (i.e., stationnarité) : les distributions à échantillon fini des statistiques usuelles des tests sont compliquées et dépendent d'un grand nombre de paramètres de nuisance inconnus. En outre, la présence de variables non stationnaires – tels que les processus intégrés

– peut affecter la distribution asymptotique des statistiques des tests et engendrer de nouveaux problèmes de fiabilité; voir par exemple, Sims, Stock et Watson (1990), Johansen (1995), Hatanaka (1996), Tanaka (1996), Dhrymes (1998), Hansen et Johansen (1998), Maddala et Kim (1998) et McAleer et Oxley (1999). En particulier, la distribution asymptotique appropriée peut dépendre de caractéristiques inconnues du processus (i.e., l'ordre d'intégration ou le nombre de relations cointégrées). C'est le cas, par exemple, dans les tests de causalité, voir Sims, Stock et Watson (1990) et Toda et Phillips (1993, 1994).

En vue de remédier à la non fiabilité des distributions asymptotiques lors de l'inférence dans les modèles VAR, des techniques de rééchantillonnage, connues aussi sous le nom de bootstrap [voir Efron et Tibshirani (1993) et Hall (1992)] ont été proposées; voir, par exemple, Jeong et Maddala (1993), Li et Maddala (1996), Paparoditis (1996), Kilian (1998b, 1998a), Caner et Kilian (1999), Kilian et Demiroglu (1997, 2000), Berkowitz et Kilian (2000), Inoue et Kilian (2002a, 2002b). Les techniques de bootstrap peuvent entraîner des améliorations spectaculaires par rapport aux approximations asymptotiques standard, mais leur justification reste toutefois intrinsèquement asymptotique. De plus, il est bien reconnu que la technique du bootstrap ne réussit pas à fournir des tests asymptotiques valides lorsque la statistique du test simulée possède une distribution asymptotique renfermant des paramètres de nuisance, en particulier lorsque cette distribution asymptotique présente des discontinuités par rapport à ces paramètres de nuisance; voir Athreya (1987), Basawa, Mallik, McCormick, Reeves et Taylor (1991), Sriram (1994), Andrews (2000), Benkwitz, Lütkepohl et Neumann (2000), et Inoue et Kilian (2002a, 2003). Ce type de situation peut facilement être présent dans les modèles VAR.

Notons que la modélisation des séries chronologiques multivariées basée sur les modèles VAR a reçu une attention considérable en économétrie; voir Lütkepohl (1991, 2001, 2005), Hamilton (1994, Chapitre 11) et Dhrymes (1998). Cependant, ces modèles requièrent souvent un grand nombre de paramètres en vue d'obtenir de meilleurs ajustements. De plus, la spécification VAR n'est pas invariante à plusieurs transformations linéaires de base. Par exemple, au lieu de satisfaire un modèle VAR,

des sous-vecteurs du modèle VAR initial, suivent plutôt des processus VARMA. Aussi, l'agrégation temporelle et contemporaine résulte en des modèles VARMA mixtes [voir Lütkepohl (1987)]. De même, l'ajustement pour la tendance et la saisonnalité peut aboutir à des modèles en dehors de la classe VAR [Maravall (1993)].

La structure VARMA inclut les modèles VAR comme cas spécial, et peut reproduire d'une manière plus parcimonieuse une classe plus grande d'autocovariances et de processus générateurs de données (PGD). Donc, les modèles VARMA peuvent conduire à une amélioration de l'estimation et de la prévision. La modélisation des modèles VARMA a été proposée il y a longtemps [voir Hillmer et Tiao (1979), Tiao et Box (1981), Lütkepohl (1991), Boudjellaba, Dufour et Roy (1992, 1994), Reinsel (1993, 1997)], mais a reçu peu d'attention en pratique. Bien que la construction des modèles VARMA reste similaire à la procédure associée au cas univarié, dans le cas multivarié, la tâche se complique davantage suite à la nature multidimensionnelle des données.

Plusieurs procédures permettant une paramétrisation unique ont été proposées au niveau de la spécification des modèles ; voir Hannan (1969b, 1970, 1971, 1976b, 1979, 1980a, 1981), Deistler et Hannan (1981), Deistler (1983), Hannan et Deistler (1988, Chapitre 2), Lütkepohl (1991, Chapitre 7) et Reinsel (1997, Chapitre 3). En vue de garantir à la fois efficacité et paramétrisation parcimonieuse, plusieurs méthodes ont été envisagées. Parmi celles-ci on peut citer : (1) les techniques basées sur l'analyse canonique [Akaike (1974a, 1975, 1976), Cooper et Wood (1982), Tiao et Tsay (1985, 1989), Tsay et Tiao (1985), Tsay (1989a) et Paparoditis et Streitberg (1991)] ; (2) l'approche basée sur l'indice de Kronecker qui spécifie une représentation en forme échelon des modèles VARMA [Deistler et Hannan (1981), Hannan et Kavalieris (1984b), Solo (1986), Tsay (1989b), Nsiri et Roy (1992, 1996), Poskitt (1987, 1992), Lütkepohl et Poskitt (1996) et Bartel et Lütkepohl (1998)] ; (3) l'approche basée sur la composante scalaire du modèle (CSM) [Tiao et Tsay (1989) et Tsay (1989b, 1991)]. Cependant, en pratique les techniques CSM et de corrélation canonique sont complexes et peuvent induire des difficultés de calcul car elles impliquent souvent l'évaluation d'un grand nombre de valeurs propres. De plus, tel qu'il a été souligné par Lütkepohl et Poskitt

(1996), les résultas de convergence sur l'estimation des indices de Kronecker moyennant l'approche basée sur la corrélation canonique ne semblent pas être disponibles. C'est pourquoi, dans ce qui suit, nous allons utiliser des critères d'information pour fin de spécification des modèles.

Une fois qu'une spécification identifiable a été formulée, des différentes méthodes d'estimation ont été considérées. Des méthodes basées sur les transformations de Fourier, le maximum de vraisemblance récursif (MV) et M-estimateurs ont été proposées respectivement par Hannan (1969a, 1980b) et Kreiss (1985, 1987). Mais la méthode d'estimation la plus largement étudiée est sans doute le MV avec erreurs Gaussiennes indépendantes et identiquement distribuées (i.i.d.) ; voir Newbold (1974), Box et Jenkins (1976), Hillmer et Tiao (1979), Nicholls et Hall (1979, 1980), Hannan, Kavalieris et Mackisack (1986), Kohn (1981), Tiao et Box (1981), Solo (1984), Shea (1989), Mauricio (2002), et l'étude de Mélard, Roy et Saidi (2002). Toutefois, la maximisation de la vraisemblance exacte dans les modèles VARMA stationnaires et inversibles requiert des calculs très lourds puisque pour chaque ordre autorégressif et moyenne mobile (p et q) une maximisation non quadratique satisfaisant des contraintes d'inégalité doit être réalisée à l'aide d'algorithmes itératifs. Tiao et Box (1981) ont souligné qu'il est beaucoup plus facile de maximiser la vraisemblance conditionnelle, bien que dans des systèmes à dimension élevée des problèmes d'ordre numérique persistent encore en raison du manque de valeurs initiales, même avec ordres (p, q) connus. De plus, les estimateurs obtenus par la méthode du quasi-maximum de vraisemblance avec des bruits faibles peuvent ne pas être convergents.

Afin de rendre la modélisation des modèles VARMA pratique, il faut avoir des méthodes d'estimation qui sont à la fois rapides et simples à mettre en œuvre avec les logiciels standard. Aussi, l'une des raisons mettant l'accent sur ces méthodes d'estimation est que la théorie distributionnelle asymptotique a tendance à être très peu fiable dans les modèles dynamiques à dimension élevée, de sorte que les tests et régions de confiance fondées sur des approximations asymptotiques sont aussi peu fiables. Ce qui laisse penser que les procédures fondées sur la simulation - par exemple, les techniques de bootstrap - devront être utilisées. Cependant, la simulation peut ne pas

être pratique si le calcul de l'estimateur est difficile ou très long.

Dans le cas univarié, Hannan et Rissanen (1982) ont proposé une méthode récursive qui exige seulement des régressions linéaires ; voir aussi Durbin (1960), Hannan et Kavalieris (1984a), Zhao-Guo (1985), Hannan, Kavalieris et Mackisack (1986), Poskitt (1987), Koreisha et Pukkila (1990a, 1990b, 1995), Pukkila, Koreisha et Kallinen (1990), Allende et Heiler (1992), Galbraith et Zinde-Walsh (1994, 1997) et Kavalieris, Hannan et Salau (2003). Cette approche est basée sur l'estimation (par moindres carrés ordinaires) des innovations du processus moyennant une longue autorégression ; ensuite les paramètres du modèle VARMA sont estimés en utilisant les résidus de la longue autorégression comme régresseurs. Par la suite, de nouveaux résidus sont filtrés et une régression linéaire utilisant des variables transformées est effectuée en vue d'atteindre l'efficacité asymptotique dans le cas d'innovations Gaussiennes.

Ces méthodes ont été étendues aux modèles VARMA ; voir Hannan et Kavalieris (1984b, 1986), Hannan et Deistler (1988), Koreisha et Pukkila (1989), Huang et Guo (1990), Reinsel, Basu et Yap (1992), Poskitt (1992), Poskitt et Lütkepohl (1995), Lütkepohl et Poskitt (1996), Lütkepohl et Claessen (1997) et De Frutos et Serrano (2002). Il est à noter que cette méthode d'estimation linéaire (dans ses deux premières étapes) a été introduite pour fin de sélection de modèles et aussi pour obtenir des valeurs initiales. Après quoi, d'autres méthodes d'estimation, comme le MV, est généralement proposée.

L'un des principaux sujets de la recherche actuelle est de développer des méthodes d'inférence à distance finie, fiables dans les modèles VAR et VARMA. Contrairement aux modèles VAR qui ne posent pas de problèmes au niveau de l'estimation et où il y a lieu seulement de montrer comment une procédure d'inférence simulée à échantillon fini, fiable, peut être mise en oeuvre, les modèles VARMA, quant à leur non linéarité, requièrent de nouvelles techniques d'estimation simples afin de rendre l'entreprise ou la conception de ces nouvelles techniques d'inférence simulée (que nous proposons pour les VAR) possible dans le cadre des modèles VARMA. Cette thèse est composée de trois essais. Le premier essai est consacré au développement de méthodes paramétriques d'inférence simulée à distance finie dans les modèles VAR, basées sur

la technique des tests de MC. Alors que les deuxième et troisième essais proposent de nouvelles techniques d'estimation simples dans les modèles VARMA, susceptibles d'être utilisées ultérieurement dans le cadre des nouvelles méthodes d'inférence proposées.

Dans le premier essai de cette thèse, nous proposons une technique d'inférence simulée à distance finie applicable à des modèles VAR paramétriques avec ordres connus. Ce qui permet de contrôler complètement le niveau du test, en dépit de la présence des paramètres de nuisance et sans hypothèses additionnelles sur la structure du processus [telles que la stationnarité et l'ordre de cointégration]. Les innovations du VAR peuvent avoir n'importe quelle distribution qui peut être spécifiée à une transformation linéaire [ou matrice de covariance] près, ce qui laisse supposer des distributions Gaussiennes ou non Gaussiennes. L'hypothèse centrale est que le modèle peut être simulé, une fois un nombre fini de paramètres a été spécifié. La technique proposée est basée sur une extension de la vieille technique des tests de Monte Carlo (MC) [Dwass (1957), Barnard (1963), Birnbaum (1974)], que nous appelons tests de *Monte Carlo maximisés* (MCM) [Dufour (2006)]. Cette méthode consiste à maximiser une fonction de p-value simulée sur l'espace des paramètres de nuisance. Deux variantes principales de cette méthode ont été considérées : la première maximise la fonction de p-value simulée sur tout l'espace des paramètres de nuisance et fournit des tests exacts sur des restrictions générales des paramètres du modèle, alors que la seconde variante limite la maximisation de la fonction de p-value simulée sur un *estimateur ensembliste convergent* des paramètres de nuisance. Cette dernière peut être vue comme une version simplifiée et approximative de la procédure exacte (parce que l'ensemble sur lequel la fonction de p-value est maximisée peut être beaucoup plus petit) ; elle fournit des tests valides asymptotiquement sans à avoir besoin d'établir la distribution asymptotique de la statistique du test ou de considérer des hypothèses additionnelles sur le structure du processus. Nous avons aussi considéré des tests de *Monte Carlo locaux* (MCL) qui peuvent être vus comme une version dégénérée de la procédure de maximisation simplifiée, obtenue en remplaçant l'estimateur ensembliste convergent par un estimateur ponctuel convergent, et peuvent être interprétés comme

des tests du bootstrap paramétrique. Bien sure, cette dernière procédure n'est pas exacte à échantillon fini et requiert des hypothèses fortes (pour permettre des tests valides asymptotiquement) comparée à la procédure de MC basée sur l'estimateur ensembliste.

Il est évident que la méthode que nous proposons repose sur des calculs intensifs nécessitant l'utilisation d'ordinateurs puissants, et l'une des contributions importantes de cet essai est de montrer que l'approche théorique proposée peut être vraiment mise en oeuvre dans un cadre dimensionnel élevé tel est le modèle VAR. Pour cette fin, nous nous sommes concentrés sur les tests du quotient de vraisemblance pour deux catégories d'hypothèses : (1) l'ordre du modèle VAR; (2) la non causalité au sens de Granger. Nous présentons des simulations sur des tests de non causalité au sens de Granger faisant évidence et documentant trois points. Le premier, est que les tests standard basés sur les valeurs critiques asymptotiques peuvent avoir des distorsions de niveau catastrophiques. Le deuxième, est que l'approche MCL (ou bootstrap) fournit des améliorations de ce point de vue, mais présente encore des fréquences notables de surrejet. Le troisième, est que dans les mêmes circonstances, l'approche MCM contrôle parfaitement le niveau du test, même lorsqu'on utilise un estimateur ensembliste des paramètres de nuisance, et de plus, présente une bonne puissance. En d'autres termes, la maximisation opérée par l'approche MCM permet des corrections effectives lorsque le bootstrap échoue (à échantillon fini et asymptotiquement). Nous appliquons aussi la méthode proposée pour tester la causalité dans un modèle trimestriel de l'économie américaine, basé sur des données utilisées dans Bernanke et Mihov (1998) et Dufour, Pelletier et Renault (2006), impliquant les réserves non empruntées, le taux des fonds fédéraux, le produit intérieur brut (PIB) réel, et le déflateur du PIB.

Dans le deuxième essai de cette thèse, nous étudions le problème d'estimation des paramètres des modèles VARMA sous la forme échelon en utilisant seulement la méthode des moindres carrés linéaires. La forme échelon a été retenue car elle tend à fournir des paramétrisations relativement parcimonieuses. En particulier, nous étudions un estimateur simple de deux étapes qui peut être obtenu au moyen de régression linéaire sur chaque équation du système VARMA, donc il s'agit bien d'un

estimateur remarquablement simple à appliquer. Cet estimateur a été déjà considéré dans les travaux mentionnés ci-dessus sur l'estimation linéaire des modèles VARMA, mais sa distribution asymptotique n'a pas été apparemment établie. Pour des indices de Kronecker donnés, caractérisant le processus VARMA, nous dérivons la distribution asymptotique de cet estimateur sous des conditions de régularité standard. En particulier, nous montrons que ce dernier possède une distribution asymptotique normale (ce qui implique sa convergence), et nous fournissons un estimateur convergent, simple, de sa matrice de covariance asymptotique, de sorte que des tests et intervalles de confiance asymptotiquement valides peuvent être construits pour les paramètres du modèle.

Dans le troisième essai, nous proposons une procédure d'estimation linéaire de trois étapes pour des modèles VARMA en forme échelon, stationnaires et inversibles. Cette approche peut être facilement adaptée aux modèles VARMAX et étendue aux modèles VARMA intégrés et cointégrés. La méthode d'estimation considère la forme échelon, puisque celle-ci tend à fournir des paramétrisations relativement parcimonieuses. Mais notre procédure reste applicable à d'autres procédures d'identification telles que celle d'équations finales ou d'autres modèles contraints pour fin d'inférence.

À la différence des travaux précédents sur les méthodes d'estimation linéaire des modèles ARMA et VARMA [à l'exception de Bartel et Lütkepohl (1998)], nous considérons la présence de constantes. De plus, nous fournissons une forme standard des paramètres estimés, ce qui est beaucoup plus facile à appliquer que celui considéré par Hannan et Kavalieris (1984b). Pour ce faire, nous étendons les résultats développés dans le deuxième essai à la méthode d'estimation linéaire généralisée en deux étapes, et nous dérivons la distribution asymptotique des estimateurs MCG (moindres carrés généralisés) correspondants. Nous donnons aussi une justification théorique simple de la troisième étape d'estimation proposée par Hannan et Kavalieris (1984b) – qu'ils ne donnent pas d'ailleurs – et nous montrons que ces estimations correspondent à une seule itération de l'algorithme du score, à partir des estimations de deuxième étape, prises comme valeurs initiales. De plus, nous prouvons sous des conditions générales que ces estimations ont la même distribution asymptotique que l'estimateur

MV dans le cas d'innovations Gaussiennes. Contrairement à Reinsel, Basu et Yap (1992), où les innovations sont supposées Gaussiennes et aucune forme d'identification n'est considérée, nous dérivons la distribution asymptotique de notre estimateur linéaire de troisième étape sous l'hypothèse de bruits blancs forts. Bien que notre procédure d'estimation de trois étapes est équivalente à celle de Hannan et Kavalieris (1984b), les estimations des covariances asymptotiques que nous donnons pour les estimateurs de troisième et deuxième étapes sont très simples et faciles à utiliser.

Outre les résultats sur l'estimation, nous proposons une procédure simplifiée de sélection d'ordre afin d'identifier les indices de Kronecker associés à la forme échelon des modèles VARMA. La procédure repose sur la capacité de déterminer la matrice implicite des restrictions pour tous les ensembles possibles des indices de Kronecker et pour n'importe quelle dimension du système VARMA. En fait, cette tâche est très difficile à réaliser et pourrait être une raison importante rendant l'utilisation de la forme échelon moins attrayante en pratique. Dans cet essai, nous développons un algorithme qui résout ce problème. Nous fournissons aussi un algorithme permettant de formuler les matrices implicites des restrictions correspondantes à chaque équation [voir le deuxième essai]. Ainsi, l'identification des indices dynamiques est beaucoup plus facile, et notre procédure tend à réduire la sur-paramétrisation et à améliorer la précision – par rapport aux méthodes proposées par Hannan et Kavalieris (1984b) et Poskitt (1992). Nous proposons également des critères d'information (pour les deux premières étapes) et des raccourcis qui donnent des estimations des indices de Kronecker fortement convergentes. Contrairement aux approches de Hannan et Kavalieris (1984b) et Poskitt (1992), qui semblent mal fonctionner dans l'estimation des indices de Kronecker, nous montrons par des simulations que notre procédure se comporte bien et donne des estimations plus précises que celles proposées par Hannan et Kavalieris (1984b). Pour illustrer notre méthode d'estimation, nous considérons le dépistage de la neutralité de la politique monétaire aux États-Unis à long terme. En effet, nous mesurons et identifions la politique monétaire américaine dans le cadre d'un modèle VARMA en forme échelon. Pour montrer l'avantage d'utiliser ces modèles parcimonieux par rapport aux modèles VAR nous étudions les fonctions d'impulsion ou de

réponse, implicites (FRIs) à partir d'un système de six séries temporelles macroéconomiques.

Chapter 1

Finite-sample simulation-based inference in VAR models with applications to order selection and causality testing

1. Introduction

Vector autoregressive (VAR) models are widely used for multivariate time series analysis, especially in econometrics; see Sims (1980), Lütkepohl (1993, 2001), Reinsel (1993), Hamilton (1994), Hendry (1995), Gouriéroux and Monfort (1997), Dhrymes (1998) and Clements and Hendry (2002). One reason for this popularity is that VAR models are easy to estimate and can account for relatively complex dynamic phenomena. Important features and applications based on such models include forecasting, causality analysis [in the sense of Wiener (1956) and Granger (1969)], impulse responses, cointegration, etc.

VAR models, however, typically involve large numbers of parameters, so that the usual statistical difficulties associated with dynamic models are compounded by high dimensionality. Not surprisingly, statistical inference (tests and confidence sets) in such models is almost universally based on large-sample approximations. Even in static multivariate linear regression models, it is well-known that asymptotic approximations can be very unreliable; see Dufour and Khalaf (2002). These problems get worse in VAR models (which can be interpreted as dynamic multivariate linear regressions), even if strong regularity assumptions (e.g., stationarity) are made : finite-sample distributions of usual test statistics are complicated and depend on large numbers of unknown nuisance parameters. Further, the presence of non-stationary variables – such as integrated processes – can affect the asymptotic distributions and lead to further reliability problems; see, for example, Sims, Stock, and Watson (1990), Johansen (1995), Hatanaka (1996), Tanaka (1996), Dhrymes (1998), Hansen and Johansen (1998), Maddala and Kim (1998), and McAleer and Oxley (1999). In particular, the appropriate asymptotic distribution may depend on unknown features of the process (e.g., the integration order or the number of cointegrating relationships). This is the case, for example, in causality testing; see Sims, Stock, and Watson (1990) and Toda and Phillips (1993, 1994).

In view of alleviating the unreliability of asymptotic distributions for inference in VAR models, bootstrap techniques [see Efron and Tibshirani (1993) and Hall (1992)]

have also been proposed; see, for example, Jeong and Maddala (1993), Li and Maddala (1996), Paparoditis (1996), Kilian (1998b, 1998a), Caner and Kilian (1999), Kilian and Demiroglu (1997, 2000), Berkowitz and Kilian (2000), Inoue and Kilian (2002a, 2002b). Bootstrap methods can lead to spectacular improvements over standard asymptotic approximations, but their justification remains intrinsically asymptotic. Further, it is well known that bootstrapping can fail to provide asymptotically valid tests when the simulated test statistic has an asymptotic distribution involving nuisance parameters, especially if the asymptotic distribution has discontinuities with respect to the nuisance parameters; see Athreya (1987), Basawa, Mallik, McCormick, Reeves, and Taylor (1991), Sriram (1994), Andrews (2000), Benkwitz, Lütkepohl, and Neumann (2000), and Inoue and Kilian (2002a, 2003). This type of situation can easily occur in VAR models.

In this paper, we propose a finite-sample simulated-based inference technique applicable to parametric finite-order VAR models that allows one to control completely the level of the test, despite the presence of large numbers of nuisance parameters and without further assumptions on the structure of the process [such as stationarity or the order of integration]. The disturbances in the VAR model may follow any distribution that is specified up to a linear transformation [or covariance matrix], which allows for both Gaussian and non-Gaussian distributions. The central assumption is that the model can be simulated once a finite number of parameters have been specified. The technique proposed is based on an extension of the old technique of Monte Carlo (MC) tests [Dwass (1957), Barnard (1963), Birnbaum (1974)], which we call *maximized Monte Carlo* (MMC) tests [Dufour (2006)]. This method involves maximizing a simulated p-value function over the nuisance parameter space. Two main variants of this method are considered : the first one maximizes the simulated p-value function over the full nuisance parameter space and yields provably exact tests of general restrictions on model parameters, while the second variant considers a maximization restricted to a *consistent set estimator* of the nuisance parameters. The latter can be viewed as an approximate simplified version of the fully exact procedure (because the set over which the p-value function is maximized can be much smaller);

it provides asymptotically valid tests without the need to establish the asymptotic distribution of the test statistic or to make further assumptions on the structure of the process. We also consider *local Monte Carlo* (LMC) tests which can be viewed as a degenerate version of the simplified maximized procedure, obtained by replacing the consistent set estimator by a consistent point estimate, and may be interpreted as parametric bootstrap test. Of course, the latter procedure is not exact in finite samples and requires stronger assumptions (to yield asymptotically valid tests) than the consistent set MC procedure.

The method proposed is obviously computer intensive, and an important contribution of this paper consists in showing that the proposed theoretical approach can indeed be implemented in a high-dimensional setup, such as a VAR model. For that purpose, we focus on likelihood ratio (LR) tests for two categories of hypotheses : (1) the order of a VAR model ; (2) Granger non-causality. We present simulation evidence on tests for Granger non-causality which document three things. First, standard tests based on asymptotic critical values can have catastrophic size properties. Second, the LMC approach (or bootstrapping) does provide improvements from that viewpoint, but can still allow for sizeable overrejection rates. Third, under the same circumstances, the MMC approach does control level perfectly, even we only use a consistent set estimator for the nuisance parameters, and provides good power. In other words, the maximization operated by the MMC approach yields effective corrections for possible failures of the bootstrap (both in finite samples and asymptotically). We also apply the proposed method to causality testing in a quarterly model of the U.S. economy, based on data previously studied in Bernanke and Mihov (1998) and Dufour, Pelletier, and Renault (2006), involving nonborrowed reserves, the federal funds rate, real gross domestic product, and the GDP deflator.

The paper is organized as follows. Section **2** describes the model and the main test problems that will be studied. Section **3** presents the principles of MC tests and MMC tests. In section **4**, we discuss how such procedures can be applied to LR-type tests in VAR models. The results of our simulation study are presented in section **5** and the empirical macroeconomic application in section **6**. We conclude in section **7**.

2. Framework

In this paper we consider a general k-dimensional finite-order vector autoregressive (VAR) process $\{Y_t : t \in \mathbb{Z}\}$ of the form :

$$Y_t = \mu + \sum_{i=1}^{p} \Phi_i Y_{t-i} + u_t, \quad t = 1, \ldots, T, \tag{2.1}$$

$$u_t = R\varepsilon_t, \quad t = 1, \ldots, T, \tag{2.2}$$

where the vectors $Y_t = (Y_{1t}, \ldots, Y_{kt})'$, $t = -p+1, \ldots, T$, are observable, p is a specified nonnegative integer ($p \geq 1$), $\mu = (\mu_1, \ldots, \mu_k)'$ is an unknown $k \times 1$ vector of intercept terms, $\Phi_i = [\Phi_{ijl}]_{j,l=1,\ldots,k}$ is an unknown $k \times k$ matrix of fixed coefficient matrices ($1 \leq i \leq p$), R is an *unknown* fixed non-singular matrix, and the conditional distribution of the vector $\varepsilon(T) = \text{vec}(\varepsilon_1, \ldots, \varepsilon_T)$, given the initial values $Y_0, \ldots, Y_{-p+1}$, is completely specified. A common special case here would consist in assuming that

$$\varepsilon_t \overset{i.i.d.}{\sim} N[0\,,\, I_k], \quad t = 1, \ldots, T, \tag{2.3}$$

given the initial values, so that the errors are independent and identically distributed (*i.i.d.*) according to a multinormal distribution $N[0\,,\, \Sigma]$ with $\Sigma = RR'$. But, from the viewpoint of implementing the procedures proposed in this paper, the assumptions (2.1) - (2.2) will be sufficient. For example, there is no need to assume normality or even the existence of moments.

Setting

$$\Phi(z) = I_k - \sum_{i=1}^{p} \Phi_i z^i, \quad z \in \mathbb{C}, \tag{2.4}$$

the model is said to be *stationary* if

$$\det\{\Phi(z)\} \neq 0 \text{ for all } |z| \leq 1, \tag{2.5}$$

and it is *stable* (non-explosive) if

$$\det\{\Phi(z)\} \neq 0 \text{ for all } |z| < 1; \tag{2.6}$$

stable models allow for the presence of roots on the unit circle for the equation $\det\{\Phi(z)\} = 0$. Note, however, that stationarity (or stability) assumptions will not be needed for the validity of the procedures proposed in this paper, so cointegrating relations may be present. The central feature we shall exploit is the fact that the model can be easily simulated, once a *finite* number of parameters have been specified.

In this paper, we consider the problem of testing general hypotheses of the form

$$H_0 : \text{vec}(\Phi) \in \Gamma_0 \tag{2.7}$$

where $\Phi = [\Phi_1, \ldots, \Phi_p]$ and $\Gamma_0 \subseteq \mathbb{R}^{k^2 p}$. This covers both linear and nonlinear hypotheses on model coefficients. In our simulations and applications, however, we shall focus on linear hypotheses, more precisely :

1. hypotheses on individual coefficients :

$$H_0(\Phi^0_{ijl}) : \Phi_{ijl} = \Phi^0_{ijl}; \tag{2.8}$$

2. hypotheses on the order of the process :

$$H_0(i) : \Phi_i = 0 \tag{2.9}$$

and

$$H_0[i\,,\, p] : \Phi_i = \cdots = \Phi_p = 0; \tag{2.10}$$

3. non-causality in the sense of Granger (1969) :

$$H_0(Y_l \nrightarrow Y_j) : \Phi_{ijl} = 0, \quad i = 1, \ldots p, \tag{2.11}$$

The distribution of most standard test statistics [such as Wald-type, LM-type or

LR-type statistics] for hypotheses on the coefficients of VAR models typically depends (under the null hypothesis) on both the hypothesis tested and unknown nuisance parameters. To be more precise, if we denote by $\mathcal{H}_0$ the set of data distributions F – or data generating processes (DGP's) – compatible with H_0, the null hypothesis can be written in the form

$$H_0 : F \in \mathcal{H}_0 \,. \tag{2.12}$$

Then a test of H_0 has *level* α iff

$$\mathsf{P}_F[\text{Rejecting } H_0] \leq \alpha \text{ for all } F \in \mathcal{H}_0 \tag{2.13}$$

or, equivalently,

$$\sup_{F \in \mathcal{H}_0} \mathsf{P}_F[\text{Rejecting } H_0] \leq \alpha, \tag{2.14}$$

and the test has *size* α iff

$$\sup_{F \in \mathcal{H}_0} \mathsf{P}_F[\text{Rejecting } H_0] = \alpha; \tag{2.15}$$

see Lehmann (1986, Chapter 3). If we also had

$$\mathsf{P}_F[\text{Rejecting } H_0] = \alpha \text{ for all } F \in \mathcal{H}_0, \tag{2.16}$$

the test would be *similar*. But, in complex models, this appears extremely difficult to achieve with any reasonable procedure that depends on the data. So we will focus on designing tests that satisfy as closely as possible the level restriction (2.13) in finite samples. So one needs a method that can adapt readily to both these features. We will now describe such a method.

3. Monte Carlo tests

In this section, we describe in a succinct way the general approach that will allow us to construct finite-sample tests for any VAR model, such as the one described in

section 2. To ensure clarity, we describe first the basic principle underlying Monte Carlo tests by considering two basic cases : (1) the distribution of the test statistic under the null hypothesis does not depend on nuisance parameters; (2) the distribution of the test statistic depends on nuisance parameters. Of course, the second case is the relevant one for inference in VAR models. To deal with it, we consider three alternative approaches : (a) *maximized Monte Carlo* (MMC) tests over the full nuisance parameter space; (b) MMC tests over a consistent set estimator of the nuisance parameters; (c) *local Monte Carlo* tests, *i.e.* Monte Carlo tests obtained after replacing the unknown nuisance parameters by a point estimate. To simplify exposition, we limit ourselves to the case where the test statistic has a continuous distribution, although it is relatively easy to extend Monte Carlo test methods to situations where the statistic follows a discrete distribution under the null hypothesis. Further details and proofs are provided in Dufour (2006) and Dufour and Khalaf (2001).

3.1. Monte Carlo tests without nuisance parameters

Let $S \equiv S(Y_1, \ldots, Y_T)$ be a test statistic for testing an hypothesis H_0, with critical region of the form :

$$S \geq c \tag{3.1}$$

where c is a constant. We will denote by S_0 the value of the test statistic based on the observed data. Suppose now that the distribution of S under H_0 does not depend on unknown nuisance parameters (and is continuous). The test has level α if

$$\mathbf{P}[S_0 \geq c] \leq \alpha \tag{3.2}$$

and size α if

$$\mathbf{P}[S_0 \geq c] = \alpha. \tag{3.3}$$

Suppose now we can generate by simulation N i.i.d. replications of S under H_0,

$$S_1, \ldots, S_N \tag{3.4}$$

independently of S_0. We can then estimate the survival function

$$G(x) = \mathsf{P}[S \geq x] \tag{3.5}$$

from the simulated samples :

$$\hat{G}_N[x;\, S(N)] = \frac{1}{N}\sum_{i=1}^{N} s(S_i - x) \tag{3.6}$$

where

$$S(N) = (S_1, \ldots, S_N)', \tag{3.7}$$

$$\begin{aligned} s(x) &= 1, \text{ if } x \geq 0, \\ &= 0, \text{ if } x < 0. \end{aligned} \tag{3.8}$$

Let us also consider

$$\hat{p}_N(x) = \frac{N\hat{G}_N[x;\, S(N)] + 1}{N+1} \tag{3.9}$$

the simulated p-value function associated with $S(N)$. Then, if N is chosen so that $\alpha(N+1)$ is an integer, it can be shown that, under H_0,

$$\mathsf{P}\left[\hat{p}_N(S_0) \leq \alpha\right] = \alpha. \tag{3.10}$$

In other words, the test which rejects H_0 when $\hat{p}(S_0) \leq \alpha$ has level α exactly.

3.2. Monte Carlo tests with nuisance parameters

We will now study the case where the distribution of the test statistic depends on nuisance parameters. In other words, we consider a model $\{(\mathcal{Z}, \mathcal{A}_{\mathcal{Z}}, \mathsf{P}_\theta) : \theta \in \Omega\}$ where we assume that the distribution of S is determined by $\mathsf{P}_{\bar{\theta}}$ [i.e., $\bar{\theta}$ is the "true" parameter vector]. We wish to test the hypothesis

$$H_0 : \bar{\theta} \in \Omega_0, \tag{3.11}$$

where $\Omega_0 \subset \Omega$. The critical region $\{S \geq c\}$, where c is a constant which does not depend of θ, has level α if and only if

$$\mathsf{P}_\theta [S \geq c] \leq \alpha, \ \forall \theta \in \Omega_0, \tag{3.12}$$

or equivalently

$$\sup_{\theta \in \Omega_0} G [S \mid \theta] \leq \alpha \tag{3.13}$$

where

$$G [x \mid \theta] = \mathsf{P}_\theta [S \geq x] . \tag{3.14}$$

Suppose now that, for each $\theta \in \Omega_0$, we generate N i.i.d. replications of S,

$$S (N, \theta) = [S_1 (\theta) , \ldots , S_N (\theta)]'$$

and compute a simulated p-value function :

$$\hat{p}_N [x \mid \theta] = \frac{N \hat{G}_N [x; S (N, \theta)] + 1}{N + 1}. \tag{3.15}$$

If $\alpha (N + 1)$ is an integer, then, under H_0,

$$\mathsf{P}\big[\sup \{\hat{p}_N [S_0 \mid \theta] : \theta \in \Omega_0\} \leq \alpha\big] \leq \alpha, \tag{3.16}$$

which means that the critical region $\sup \{\hat{p}_N [S_0 \mid \theta] : \theta \in \Omega_0\} \leq \alpha$ has level α. This procedure will be called a *maximized Monte Carlo test.* It allows one to obtain provably exact tests based on any statistic that can be simulated once a finite number of nuisance parameters have been specified.

The simulated p-value function $\hat{p}_N [S_0 \mid \theta]$ is not continuous, so standard gradient based methods cannot be used to maximize it. But search methods applicable to non-differentiable functions are applicable, e.g. *simulated annealing* [see Goffe, Ferrier, and Rogers (1994)].

3.3. MMC tests based on consistent set estimators

Suppose now that the test statistic depends on a sample of size T,

$$S = S_T\,, \tag{3.17}$$

and we have a consistent set estimator C_T of θ (under H_0) :

$$\lim_{T\to\infty} \mathsf{P}\left[\bar{\theta} \in C_T\right] = 1. \tag{3.18}$$

For example, if $\hat{\theta}_T$ is a consistent point estimate of $\bar{\theta}$ and c is any positive constant, the set

$$C_T = \{\theta \in \Omega : \| \hat{\theta}_T - \theta \| < c\} \tag{3.19}$$

is a consistent set estimator of $\bar{\theta}$:

$$\lim_{T\to\infty} \mathsf{P}\left[\bar{\theta} \in C_T\right] = \lim_{T\to\infty} \mathsf{P}\left[\|\hat{\theta}_T - \bar{\theta}\| < \varepsilon\right] = 1\,,\ \forall \varepsilon > 0. \tag{3.20}$$

Assuming that, for each $\theta \in \Omega_0$, we can generate N i.i.d. replications of S_T, say $S_{T1}(\theta), \ldots, S_{TN}(\theta)$, we have, under H_0 :

$$\lim_{T\to\infty} \mathsf{P}\left[\sup\{\hat{p}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \leq \alpha\right] \leq \alpha \tag{3.21}$$

where S_{T0} is the value of S_T based on the observed data, and

$$\hat{p}_{TN}[x \mid \theta] = \frac{N\hat{G}_N[x;\, S_T(N,\theta)] + 1}{N+1}, \quad S_T(N,\theta) = [S_{T1}(\theta), \ldots, S_{TN}(\theta)]'. \tag{3.22}$$

In other words, the critical region $\sup\{\hat{p}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \leq \alpha$ has level α. No assumption on the asymptotic distribution of S_T is needed.

An obvious alternative would consist in taking

$$C_T = \{\hat{\theta}_T\}, \tag{3.23}$$

which suggests one to use a critical region of the form

$$\hat{p}_{TN}[S_{T0} \mid \hat{\theta}_T] \leq \alpha. \tag{3.24}$$

We shall call this procedure a *local Monte Carlo* test. Under additional regularity conditions, it is possible to show that

$$\lim_{T\to\infty} \mathsf{P}\left[\hat{p}_{TN}[S_{T0} \mid \hat{\theta}_T] \leq \alpha\right] \leq \alpha \tag{3.25}$$

but the conditions under which this holds are notably more restrictive than those under which (3.21) obtains. This procedure may also be interpreted as a *parametric bootstrap*, except for the fact that the number of replications N is explicitly taken into account.

A good consistent restricted estimate $\hat{\theta}_T$ is typically a reasonable starting point for computing the MMC p-value. Since

$$\hat{p}_{TN}\left[S_{T0} \mid \hat{\theta}_T\right] \leq \sup\left\{\hat{p}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in \Omega_0\right\}, \tag{3.26}$$

it is clear that

$$\hat{p}_{TN}\left[S_{T0} \mid \hat{\theta}_T\right] > \alpha \tag{3.27}$$

implies

$$\sup\left\{\hat{p}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in \Omega_0\right\} > \alpha. \tag{3.28}$$

A non-significant bootstrap p-value entails a non-significant MMC p-value. The MMC procedure offers protection against failures of the bootstrap.

4. Tests in VAR models

We will now consider the problem of testing restrictions on the coefficients Φ of model (2.1) - (2.2). Even though various procedures, such as Wald-type, LM-type or LR-type tests, may be used, we will concentrate here on LR tests based on statistics

of the form :

$$LR = 2\left[\ln L(\hat{\delta}) - \ln L(\hat{\delta}^0)\right] \tag{4.1}$$

where $L(\cdot)$ is the likelihood function, $\hat{\delta}$ is the unconstrained maximum likelihood (ML) estimator of the model parameter vector $\delta \equiv \text{vec}[\mu, \Phi, R]$ obtained by maximizing the likelihood function over the full feasible parameter space, and $\hat{\delta}^0$ is the constrained ML estimator. Since a specific likelihood function must be specified, we shall focus on Gaussian LR statistics.

Under the assumption that the errors u_t, $t = 1, \ldots, T$, conditional on the initial values $y_{-p} = vec[Y_0, \ldots, Y_{-p+1}]$, are i.i.d. $N[0, \Sigma]$, the likelihood function is

$$L(\delta) = \kappa - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T} u_t'\Sigma^{-1}u_t \tag{4.2}$$

where κ is a constant and

$$u_t = Y_t - \mu - \sum_{i=1}^{p}\Phi_i Y_{t-i}, \quad t = 1, \ldots, T. \tag{4.3}$$

Then the (conditional) LR statistic for testing any hypothesis of the form H_0 : $\text{vec}(\Phi) \in \Gamma_0$, is

$$LR_G = T\ln\left(\Lambda_T^0\right) \tag{4.4}$$

with

$$\Lambda_T^0 = \left|\hat{\Sigma}_T^0\right| / \left|\hat{\Sigma}_T\right| \tag{4.5}$$

where $\hat{\Sigma}_T^0$ and $\hat{\Sigma}_T$ are respectively the restricted and unrestricted ML estimators of the error covariance matrix Σ. For stationary processes, under standard regularity conditions, the asymptotic distribution of the LR statistic under the null hypothesis is chi-square with number of degrees equal to the number of (linearly independent) constraints. This will be the case in particular for zero coefficient restrictions (2.8), restrictions on the order of the process (2.10) and Granger non-causality restrictions (2.11).

For example, consider the problem of testing a Granger non-causality hypothesis,

such as $H_0(Y_l \nrightarrow Y_j)$ in (2.11). Here, all the coefficients of the VAR which are not fixed by this null hypothesis – *i.e.* the unknown coefficients of μ, Φ or R – may appear as nuisance parameters in the distribution of LR_G. Further, once the nuisance parameters are set, the model (2.1) - (2.2) and the corresponding test statistic can be simulated. So we propose using Monte Carlo test procedures adapted to the presence of nuisance parameters, in particular maximized Monte Carlo tests. This means applying the MMC procedures described in section 3 with $S(\theta)$ replaced by LR_G, where θ may stand for the elements of δ which are fixed by the null hypothesis.

Such procedures are obviously computer intensive and require dynamic simulations of the process under various parameter configurations (compatible with the null hypothesis). Explosive parameter are not necessarily excluded by the estimation procedure or the model considered. But parameter values can lead to numerical problems (overflows), so one may wish to exclude explosive processes. In VAR models, such restrictions may not be easy to impose. For that purpose, it is useful to note that the roots of the determinantal equation $\det[\Phi(z)] = 0$ are identical with the inverses of the eigenvalues of the matrix

$$\bar{\Phi} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ I_k & 0 & \cdots & 0 & 0 \\ 0 & I_k & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & I_k & 0 \end{bmatrix}. \tag{4.6}$$

The corresponding VAR(p) process is stationary when these eigenvalues are all *inside* the unit circle [see Lütkepohl (1993, Chapter 2) and, for a proof, Appendix A below]. Given the availability of efficient algorithms for computing eigenvalues, this can provide a useful way of excluding explosive processes or limiting the degree of explosion.

The algorithm for computing the MC p-values can be described as follows :

1. choose the restricted subset of the parameter space Ω_0 over which the maxi-

mization required by the MMC procedure will be performed; Ω_0 may be the whole parameter space restricted by the null hypothesis (and, eventually, by stationarity or stability restrictions) or a consistent restricted set estimator;

2. compute the test statistic $LR^{(0)}$ based on the observed data;
3. generate i.i.d. error vectors $\varepsilon^{(l)} = \left[\varepsilon_1^{(l)}, \ldots, \varepsilon_T^{(l)}\right]$, $l = 1, \ldots, N$, according to the selected distribution – for example, $\varepsilon_t^{(l)} \overset{i.i.d.}{\sim} N[0, I_k]$, $t = 1, \ldots, T$ – and construct pseudo-samples as functions of the model parameter vector $\delta = \text{vec}[\mu, \Phi, R]$:

$$Y_t^{(l)} = \mu + \sum_{i=1}^{p} \Phi_i Y_{t-i}^{(l)} + R\varepsilon_t^{(l)}, \quad t = 1, \ldots, T, \; l = 1, \ldots, N; \tag{4.7}$$

4. compute the corresponding test statistics as $LR^{(l)}(\delta)$, $l = 1, \ldots, N$, which in turn can be viewed as functions of δ and $\varepsilon^{(l)}$;
5. compute the simulated p-value function

$$\hat{p}_N[x \mid \delta] = \frac{N\hat{G}_N[x \mid \delta] + 1}{N+1}, \quad \hat{G}_N[x \mid \delta] = \frac{1}{N}\sum_{l=1}^{N} s\big(LR^{(l)}(\delta) - x\big); \tag{4.8}$$

6. compute the LMC p-value $\hat{p}_N\big[LR^{(0)} \mid \hat{\delta}_T^0\big]$ where $\hat{\delta}_T^0$ is the constrained estimator based on the observed data; if $\hat{p}_N\big[LR^{(0)} \mid \hat{\delta}_T^0\big] > \alpha$, the LMC test is not significant at level α and it is clear the MMC test does not reject either at level α (so the process can be stopped).

The procedure just described can be interpreted as the generation of a parametric bootstrap p-value. Of course, to the extent that the point estimate is typically different from the "true" parameter, the test so obtained is not exact. The MMC procedure involves maximizing the p-value function over the nuisance parameter space, as follows :

1. compute the maximized p-value

$$\hat{p}_{MMC} = \sup\left\{\hat{p}_N\big[LR^{(0)} \mid \delta\big] : \delta \in \Omega_0\right\}; \tag{4.9}$$

2. reject the null hypothesis at level α if $\hat{p}_{MMC} \leq \alpha$.

When evaluating $\hat{p}_{MMC}$, it is important to note that δ is the only free variable; the observed test statistic and the simulated errors $\varepsilon^{(l)}$, $l = 1, \ldots, N$, are treated as fixed. Even if the LMC test procedure is not significant, it may still be of interest to compute $\hat{p}_{MMC}$ to get a better idea how strongly the null is accepted. As indicated above, the maximization yields a procedure such that the probability of rejection under the null hypothesis is not larger than the level, irrespective of the unknown value of the nuisance parameters. In practice, a reasonable strategy would consist in maximizing the p-value function by taking $\delta = \hat{\delta}_T^0$ as the starting value : even if the maximization is not complete, this provides an immediate safeguard against bootstrap p-values that would be highly sensitive to nuisance parameters. As described in section 3.3, if the region over which we maximize is properly designed (as a consistent set estimator), this yields an asymptotically valid test even if the parametric bootstrap test does not.

5. Simulation experiment

In this section, we present simulation evidence on the performance of three basic types of procedures for hypothesis testing in VAR models : (1) standard tests based on asymptotic chi-square critical values; (2) local Monte Carlo tests (or parametric bootstrap tests), based on a single consistent restricted estimator of model nuisance parameters; (3) maximized Monte Carlo tests. In view of allowing for VAR processes which are non-stationary (integrated) or with roots close to unit circle, we also consider lag-augmented Wald tests proposed by Dolado and Lütkepohl (1996), Toda and Yamamoto (1995), Yamada and Toda (1998) and Dufour, Pelletier, and Renault (2006).[1] The latter procedure have the feature of leading to usual normal (or chi-square) asymptotic distributions, even when the process is integrated, so that we can expect smaller size distortions. Below, we shall consider parametric and non-parametric bootstrap versions of this procedure. A detailed description of the way

[1]For related results, see also Sims, Stock, and Watson (1990), Park and Phillips (1989), Choi (1993), Yamamoto (1996), and Kurozumi and Yamamoto (2000).

lag-augmented Wald tests were implemented in this study is presented in Appendix B.[2]

For the purpose of this experiment, we considered standard VAR(p) models with Gaussian disturbances :

$$Y_t = \sum_{i=1}^{p} \Phi_i Y_{t-i} + u_t\,, \quad t = 1, \ldots, T, \tag{5.1}$$

$$u_t = R\varepsilon_t\,, \quad t = 1, \ldots, T, \tag{5.2}$$

$$\varepsilon_t \overset{i.i.d.}{\sim} N[0\,,\, I_k]\,, \quad t = 1, \ldots, T. \tag{5.3}$$

The null hypothesis tested is Granger non-causality

$$H_0 : (Y_2, \ldots, Y_k) \nrightarrow Y_1 \tag{5.4}$$

which is equivalent to

$$H_0 : \Phi_{i1l} = 0\,,\ i = 1, \ldots, p\,,\ l = 2\,, \ldots, k. \tag{5.5}$$

Various dimensions ($k = 2, \ldots, 6$), autoregressive orders ($p = 1, \ldots, 5$), sample sizes ($T = 30, 50, 100, 200, 300$) and parameter structures (Φ) were considered. Under the null hypothesis, the data generating processes have the following relatively simple structure :

$$\Phi(L) = (1 - \varphi L)^p \otimes I_k \tag{5.6}$$

where φ is scalar which determines the degree of persistence in the series. Clearly, the process is stationary when $|\varphi| < 1$. R is a nonsingular lower triangular matrix (the values of R used in this experiment are given in Appendix C). The nominal level of the tests is 0.05. The test statistic considered is the LR-type statistic described in section 4. Monte Carlo tests (local and maximized) are based on $N = 99$ replications for tables i, v and vi, $N = 999$ for tables ii - iv, while the number of trials used for

[2]We are grateful to a referee for suggesting that we study such a method in the context of our simulation.

evaluating rejection frequencies is 1000.[3] Local Monte Carlo tests are based on the restricted ML estimator, while the MMC tests are based on maximizing the p-value in a box obtained by taking 5 units on each side of the restricted ML estimator. Some of the results of our experiment are presented in tables i to v (rejection proportions are expressed in percentages). In Table i, models M0, M1, M2, M3, M4 are based respectively on the following values of the persistence parameter : $\varphi = 0.9$, 0.95, 0.99, $(0.95)^{kp}$, $(0.99)^{kp}$. In Table v, models M2 (panel A) and M0 (panels B, C) are the basic models used, except for modifications to Φ_{i1l} in order to evaluate power. Namely, power is assessed by changing the values of the coefficients Φ_{i1l}, $i = 1, \ldots, p$, as follows : $\Phi_{i1l} = \bar{\Phi}_{i1l}(p) \neq 0$, $l = 2, \ldots, k$, where $\bar{\Phi}_{i1l}(p)$ depends on the order p of the process : $\bar{\Phi}_{i1l}(1) = 0.1$, $\bar{\Phi}_{i1l}(2) = 0.02$, and $\bar{\Phi}_{i1l}(p) = 0.01$ for $p = 3$, 4, 5. Initial values were set equal to zero. The simulations were all run with GAUSS.

From the results in tables i and v, we see clearly that asymptotic tests based on standard chi-square critical values can have catastrophic size properties, with rejection frequencies as high as 0.97 (instead of 0.05). Using local Monte Carlo (or bootstrap) test does provide important improvements, but overrejections can still be much higher than the level (for example, 0.59 rather than 0.05). Interestingly, the lag-augmented Wald tests can also severely over-reject, even when they are implemented using bootstrap methods (tables ii - iv). By contrast, in Table v, we can see that the MMC procedure controls very well the level of the test, allows and provides good power under the alternative. Indeed, it is the only method that allows one to do that.[4] Finally, Table vi shows that the MMC has power which behaves in a normal way, increasing as the model moves away from the null hypothesis.

[3] This relatively small number was used because the restricted model requires a nonlinear estimation and the simulation-based tests are themselves computer intensive.

[4] We do not report power evaluations for the asymptotic and bootstrap tests, because the level of these procedures cannot be controlled in practice.

Table i. Empirical levels of Granger causality tests with nominal level $\alpha = 0.05$

(A) VAR(1) models with different dimensions $k = 2, 3, 4, 5, 6$; $T = 30$

	Model M0		Model M1		Model M2	
k	ASY_{lr}	LMC_{lr}^{pa}	ASY_{lr}	LMC_{lr}^{pa}	ASY_{lr}	LMC_{lr}^{pa}
2	9.9	6.1	12.7	7.1	15.3	7.7
3	13.4	6.5	16.3	7.6	19.5	8.6
4	17.7	7.1	21.3	8.7	26.2	9.3
5	21.8	7.9	25.2	9.4	29.6	10.7
6	26.3	8.8	32.4	10.2	35.1	12.1

(B) Bivariate VAR(p) models, $p = 1, 2, 3, 4, 5$; $T = 30$

	Model M0		Model M3		Model M4	
p	ASY_{lr}	LMC_{lr}^{pa}	ASY_{lr}	LMC_{lr}^{pa}	ASY_{lr}	LMC_{lr}^{pa}
1	10.0	5.3	14.1	7.6	16.3	7.8
2	25.9	10.6	28.0	10.4	32.5	10.2
3	44.9	17.9	39.8	13.3	50.1	18.0
4	64.8	26.5	47.8	14.7	64.5	25.9
5	76.7	36.4	60.0	16.8	75.3	30.8

(C) VAR models [(5.6) with $\varphi = 0.90$]

	k	2		3		4		5	
T	p	ASY_{lr}	LMC_{lr}^{pa}	ASY_{lr}	LMC_{lr}^{pa}	ASY_{lr}	LMC_{lr}^{pa}	ASY_{lr}	LMC_{lr}^{pa}
50	1	7.9	4.8	10.6	5.9	13.1	7.8	16.5	7.2
	2	16.8	7.8	30.6	9.8	45.8	13.6	60.5	18.6
	3	35.3	11.8	62.2	19.8	81.7	26.2	91.7	37.9
	4	57.0	18.9	83.2	34.1	93.5	49.5	97.0	59.0
	5	69.4	26.5	91.0	49.8	96.9	58.2	97.7	59.3
100	1	7.2	5.2	9.5	5.9	10.3	6.7	10.5	4.9
	2	11.0	5.9	16.7	7.2	24.6	8.0	36.7	10.9
	3	20.2	8.8	34.9	10.0	60.6	14.1	78.0	20.3
	4	34.4	10.2	68.7	16.6	88.8	33.5	96.8	48.6
	5	53.4	15.0	87.2	33.5	98.8	45.4	98.1	50.2
200	1	5.7	5.5	6.0	4.3	7.2	5.1	7.5	5.1
	2	8.9	5.7	10.5	6.3	14.4	7.4	17.8	6.0
	3	11.7	5.9	18.8	7.2	26.5	7.5	42.6	10.7
	4	16.2	6.8	30.6	8.8	54.4	11.5	73.9	18.2
	5	24.5	9.0	51.4	12.0	82.6	19.1	96.0	25.0
300	1	5.5	4.4	6.5	5.4	6.6	4.6	8.3	6.1
	2	7.4	5.5	9.0	6.3	9.3	4.3	14.0	6.7
	3	8.3	4.8	13.1	5.7	18.0	7.3	23.9	7.5
	4	10.4	4.8	17.0	5.2	28.3	5.8	49.7	9.8
	5	13.1	5.8	27.0	7.1	55.6	11.8	78.9	16.1

Note – ASY_{lr} stays for the asymptotic test based on the likelihood ratio statistic while LMC_{lr}^{pa} is the corresponding local MC (parametric bootstrap) p-value. Models M0, M1, M2, M3 and M4 correspond to (5.6) with $\varphi = 0.90, 0.95, 0.99, (0.95)^{kp}, (0.99)^{kp}$ respectively. The proportions in this table as well subsequent tables are written in percentages.

Table ii. Causality tests based on lag-augmented Wald statistics.
Empirical levels of asymptotic, parametric bootstrap and nonparametric bootstrap procedures.($\alpha = 0.05$).
Autoregressive matrix polynomials of the form (5.6) with $\varphi = 0.90$

T	50			100			200			300		
(k, p)	ASY_{aw}	LMC_{aw}^{pa}	LMC_{aw}^{np}	ASY_{aw}	LMC_{aw}^{pa}	LMC_{aw}^{np}	ASY_{aw}	LMC_{aw}^{pa}	LMC_{aw}^{np}	ASY_{aw}	LMC_{aw}^{pa}	LMC_{aw}^{np}
(2, 1)	8.1	5.4	5.3	6.9	5.9	5.4	4.8	4.5	4.5	4.7	4.3	4.5
(2, 2)	11.6	5.0	5.2	9.0	5.7	5.5	6.6	5.0	4.8	4.7	4.1	4.1
(2, 3)	30.7	9.3	9.6	15.8	6.4	6.5	11.0	7.0	7.1	7.1	4.8	4.7
(2, 4)	58.3	16.1	15.9	30.0	9.3	9.1	11.5	4.7	4.5	9.1	5.5	5.2
(2, 5)	79.5	22.9	23.2	50.9	13.2	13.0	19.4	4.7	4.8	11.6	4.3	4.5
(3, 1)	9.9	5.7	5.5	7.5	5.3	5.9	6.5	5.4	4.8	6.2	5.6	5.9
(3, 2)	24.9	8.4	8.4	9.9	5.1	4.8	6.7	3.8	4.0	7.0	5.0	4.6
(3, 3)	63.1	13.1	13.6	32.8	8.7	8.7	14.2	5.5	5.9	10.1	5.0	4.8
(3, 4)	90.8	29.6	28.9	63.1	15.3	15.4	31.1	9.0	8.9	17.3	5.9	6.0
(3, 5)	98.4	41.3	40.4	89.6	31.4	32.0	53.4	10.4	10.4	25.6	6.1	6.2
(4, 1)	13.2	5.4	4.8	8.3	5.1	5.6	7.2	5.5	5.7	5.3	4.4	4.4
(4, 2)	40.6	9.3	9.6	19.9	6.6	6.4	9.9	4.6	5.1	7.3	4.9	4.8
(4, 3)	84.4	21.1	21.0	53.9	12.9	12.5	23.9	6.6	7.1	15.2	6.2	6.2
(4, 4)	99.0	38.5	37.4	88.6	26.7	26.4	49.5	10.4	10.4	28.3	7.1	7.1
(4, 5)	100	49.8	49.6	98.7	50.0	49.7	83.6	19.4	19.5	52.9	8.2	8.7
(5, 1)	16.2	6.1	6.4	6.8	3.7	3.5	7.6	5.9	5.8	6.6	4.8	5.0
(5, 2)	54.5	9.8	9.8	29.0	8.4	8.3	12.6	5.4	5.7	9.9	5.7	5.7
(5, 3)	95.2	25.1	24.8	73.0	16.2	17.3	37.9	7.9	7.8	22.0	5.7	5.9
(5, 4)	100	46.6	47.3	97.4	39.0	38.9	75.5	16.4	16.2	41.7	8.6	8.7
(5, 5)	100	61.4	60.9	100	65.9	65.5	96.5	33.5	33.6	76.7	14.4	14.8

Note – ASY_{aw} represents asymptotic p-values for the lag-augmented Wald statistic, while LMC_{aw}^{pa} and LMC_{aw}^{np} are the related parametric and nonparametric bootstrap p-values, respectively.

Table iii. Causality tests based on lag-augmented Wald statistics.
Empirical levels of asymptotic, parametric bootstrap and nonparametric bootstrap procedures.($\alpha = 0.05$).
Autoregressive matrix polynomials of the form (5.6) with $\bar{\varphi} = 0.95$

T	50			100			200			300		
(k, p)	ASY_{aw}	LMC^{pa}_{aw}	LMC^{np}_{aw}	ASY_{aw}	LMC^{pa}_{aw}	LMC^{np}_{aw}	ASY_{aw}	LMC^{pa}_{aw}	LMC^{np}_{aw}	ASY_{aw}	LMC^{pa}_{aw}	LMC^{np}_{aw}
(2, 1)	6.5	3.8	3.9	5.2	4.2	3.6	5.6	5.3	5.4	5.6	5.5	5.2
(2, 2)	18.3	7.2	7.0	8.9	4.8	4.8	6.4	4.2	4.3	5.9	5.1	5.2
(2, 3)	38.0	11.4	11.6	22.3	8.3	8.2	14.2	6.1	6.3	10.7	6.3	6.2
(2, 4)	70.4	24.1	24.1	51.1	18.0	17.4	25.0	8.0	8.2	15.4	7.6	6.6
(2, 5)	85.5	31.4	31.3	68.7	22.8	23.1	41.3	11.4	11.3	28.4	8.5	8.1
(3, 1)	11.9	6.6	6.1	7.0	4.8	4.8	4.8	4.1	3.8	6.5	5.6	5.6
(3, 2)	28.9	8.2	8.2	17.9	8.0	7.0	10.4	6.5	6.5	7.6	5.2	5.4
(3, 3)	72.5	19.8	19.7	48.3	15.3	15.3	24.0	6.9	7.0	16.5	7.9	7.9
(3, 4)	92.9	37.1	37.3	79.5	31.0	31.0	54.8	17.7	17.9	35.1	10.5	10.7
(3, 5)	99.0	43.7	43.5	94.9	46.4	46.2	81.9	24.2	24.0	62.5	12.1	12.5
(4, 1)	13.8	5.5	6.2	9.5	6.3	6.1	5.0	3.4	3.7	6.3	5.4	5.5
(4, 2)	47.6	10.9	10.9	24.6	6.8	7.2	14.1	6.6	6.5	10.4	5.7	5.7
(4, 3)	89.1	27.7	27.6	68.3	18.1	18.7	45.3	13.5	13.6	28.2	9.5	9.5
(4, 4)	99.1	44.0	44.6	94.7	42.6	43.3	83.5	25.6	24.5	64.2	15.1	15.3
(5, 1)	17.4	6.5	6.5	10.4	5.2	5.1	5.6	3.6	3.4	6.3	5.3	5.1
(5, 2)	60.2	12.7	12.1	37.3	10.0	9.7	19.1	6.5	6.4	15.3	6.4	6.2
(5, 3)	97.7	33.9	34.3	81.8	27.3	27.2	63.3	17.2	17.2	43.1	11.6	11.7
(5, 4)	100	56.0	55.9	98.7	56.0	55.7	94.7	38.1	37.2	82.3	22.0	21.8

Table iv. Causality tests based on lag-augmented Wald statistics.
Empirical levels of asymptotic, parametric bootstrap and nonparametric bootstrap procedures.($\alpha = 0.05$).
Autoregressive matrix polynomials of the form (5.6) with (φ, T) = (0.99, 50), (0.98, 100), (0.97, 200) and (0.96, 300)

T	50			100			200			300		
(k, p)	ASY_{aw}	LMC^{pa}_{aw}	LMC^{np}_{aw}	ASY_{aw}	LMC^{pa}_{aw}	LMC^{np}_{aw}	ASY_{aw}	LMC^{pa}_{aw}	LMC^{np}_{aw}	ASY_{aw}	LMC^{pa}_{aw}	LMC^{np}_{aw}
(2, 1)	7.9	4.4	4.5	6.8	5.5	5.2	5.5	5.2	5.3	4.9	4.5	4.8
(2, 2)	21.2	8.7	8.3	12.4	5.7	5.7	9.1	5.4	6.0	6.6	4.9	4.8
(2, 3)	48.5	16.4	16.6	36.6	14.6	15.1	18.2	6.8	7.2	12.3	6.7	6.9
(2, 4)	70.2	29.1	29.0	61.2	24.1	24.4	38.2	14.3	13.8	22.0	8.9	9.0
(2, 5)	84.4	27.4	27.8	73.8	23.6	23.8	57.0	14.2	14.4	31.7	10.4	9.7
(3, 1)	10.8	5.4	5.3	6.8	4.6	4.6	6.9	5.9	5.8	6.3	5.6	6.1
(3, 2)	34.8	11.3	11.5	22.9	9.5	9.8	11.7	4.9	5.0	9.4	6.1	6.6
(3, 3)	75.9	27.0	26.8	60.2	22.2	22.3	40.4	14.6	14.6	21.3	6.2	6.3
(3, 4)	95.2	45.5	45.2	86.5	38.1	38.5	70.8	24.0	23.2	46.7	13.4	14.0
(3, 5)	98.4	42.1	41.7	94.2	36.9	37.4	87.2	30.2	30.2	72.1	16.4	16.0
(4, 1)	15.3	6.5	6.5	8.7	5.8	5.8	7.5	6.0	6.2	6.5	5.4	5.4
(4, 2)	50.3	13.3	13.1	32.5	9.9	10.5	19.9	9.1	9.2	13.1	7.1	6.7
(4, 3)	92.0	34.9	34.6	79.9	29.5	29.3	57.0	17.5	18.3	35.8	9.4	10.2
(4, 4)	99.1	49.9	51.1	95.4	52.1	51.7	88.9	38.0	37.8	73.6	19.7	19.5
(5, 1)	17.2	5.2	5.2	9.2	4.3	4.4	8.1	5.4	5.6	6.5	4.8	5.1
(5, 2)	66.3	16.1	15.1	42.0	13.6	13.5	24.2	8.6	8.8	17.9	7.6	7.8
(5, 3)	98.3	38.4	38.7	90.2	35.7	35.6	71.9	22.6	22.1	51.7	14.4	14.4
(5, 4)	100	56.5	56.1	99.6	59.6	59.7	97.5	47.1	47.6	91.1	29.9	30.4

Table v. Empirical levels of asymptotic, LMC and MMC Granger causality tests with nominal level $\alpha = 0.05$

(A) VAR(1) models with different dimensions $k = 2, 3, 4, 5, 6$; $T = 30$

k	Level			Power
	ASY_{lr}	LMC_{lr}^{pa}	MMC_{lr}^{pa}	MMC_{lr}^{pa}
2	15.5	7.7	3.2	84.7
3	22.0	9.0	3.3	95.1
4	24.7	9.7	2.8	99.8
5	32.2	11.9	3.1	99.4
6	35.1	12.1	2.7	92.1

(B) Bivariate VAR(p) models of different orders $p = 1, 2, 3, 4, 5$; $T = 30$

p	Level			Power
	ASY_{lr}	LMC_{lr}^{pa}	MMC_{lr}^{pa}	MMC_{lr}^{pa}
1	10.5	5.1	0.4	70.0
2	25.7	8.9	0.9	56.4
3	45.2	15.8	2.3	74.2
4	64.3	25.3	4.7	85.1
5	78.0	39.2	4.2	96.2

(C) k-dimensional VAR(p) models with different sample sizes

T	50			100			200			300		
(k, p)	ASY_{lr}	LMC_{lr}^{pa}	MMC_{lr}^{pa}	ASY_{lr}	LMC_{lr}^{pa}	MMC_{lr}^{pa}	ASY_{lr}	LMC_{lr}^{pa}	MMC_{lr}^{pa}	ASY_{lr}	LMC_{lr}^{pa}	MMC_{lr}^{pa}
(2, 1)	10.2	6.5	1.6	6.4	5.3	1.1	5.9	4.9	0.8	6.4	5.2	1.0
(2, 2)	18.9	8.9	1.2	11.1	5.2	0.6	8.4	6.2	0.7	7.7	4.8	0.4
(2, 3)	36.8	11.9	2.2	19.2	6.1	0.3	10.5	5.1	1.2	7.5	4.9	0.6
(2, 4)	60.1	18.6	4.4	33.2	10.7	1.5	16.2	5.5	0.3	11.1	6.9	0.3
(2, 5)	69.2	25.9	3.8	51.8	15.3	2.7	24.8	8.2	1.2	12.4	5.6	0.8
(3, 1)	12.2	7.2	0.8	9.0	5.4	1.4	6.1	5.7	1.1	6.5	5.0	1.8
(3, 2)	29.2	9.9	1.5	18.0	7.8	2.3	10.3	5.2	1.1	8.1	5.0	0.9
(3, 3)	63.8	18.5	2.0	37.6	10.5	1.7	19.5	6.5	0.3	11.7	5.7	0.2
(3, 4)	85.2	36.5	2.2	69.3	18.9	4.1	31.2	8.6	1.2	18.3	5.9	0.4
(3, 5)	92.1	48.3	3.6	88.1	32.7	4.3	50.8	11.4	1.7	26.3	7.4	1.3
(4, 1)	15.1	6.7	1.2	9.3	6.1	1.0	7.0	4.4	1.0	6.2	5.1	1.7
(4, 2)	45.4	13.4	3.1	26.6	8.6	1.4	12.7	5.4	0.1	10.6	5.5	0.1
(4, 3)	82.0	28.6	3.3	59.4	15.5	0.2	28.7	6.7	0.1	17.2	6.4	0.1
(4, 4)	93.8	48.9	3.2	86.9	34.1	2.9	52.7	10.9	0.6	27.3	6.2	0.9
(4, 5)	97.8	58.5	4.1	98.9	46.3	3.1	84.1	18.7	2.1	57.9	11.2	1.9

Note - ASY_{lr} stays for the asymptotic test based on the likelihood ratio statistic while LMC_{lr}^{pa} and MMC_{lr}^{pa} are the corresponding local MC (parametric bootstrap) and maximized MC p-values, respectively. Panel A is based on model M2 ($\varphi = 0.99$), while panels B and C are based on model M0 ($\varphi = 0.90$). Power is obtained under alternatives where $\Phi_{iil} = \bar{\Phi}_{iil}(p) \neq 0$, $l = 2, \ldots, k$, $i = 1, \ldots, p$, where $\bar{\Phi}_{iil}(p)$ depends on the order p of the process: $\bar{\Phi}_{iil}(1) = 0.1$, $\bar{\Phi}_{iil}(2) = 0.02$, and $\bar{\Phi}_{iil}(p) = 0.01$ for $p = 3, 4, 5$.

6. Application to a VAR model of the U.S. economy

In this section, we present an application of the techniques proposed above to test Granger causality in the context of a VAR model involving four U.S. macroeconomic variables. The data used come from a study of U.S. monetary policy due to Bernanke and Mihov (1998); see also Dufour, Pelletier, and Renault (2006). This data set consists of monthly observations on nonborrowed reserves (NBR, also denoted M), the federal funds rate (FFR, r), real gross domestic product (GDP, y), the GDP deflator (GDPD, P). The monthly data on GDP and GDP deflator were constructed by state space methods from quarterly observations [see Bernanke and Mihov (1998) for more details]. The sample goes from January 1965 to December 1996. In what follows, all the variables were transformed by a logarithmic transformation. For the purpose of the study, the data were also aggregated to get quarterly observations (using arithmetic averages) and put in first differences so that we roughly consider growth rates. Results based on both quarterly and monthly data are presented below. Of course, the monthly models involves considerably more nuisance parameters. Monte Carlo tests in this example are based on $N = 999$ replications, while the MMC tests are based on maximizing the p-value in a box obtained by taking 5 units on each side of the restricted ML estimator.

The first problem we face consists in specifying the order of the VAR. Using quarterly data [see table vii], we found that the MC tests reject much less often than the asymptotic procedure : LMC tests are significant procedure at level 5% for the orders 0, 1, 2, 3, 8, 9 [plus 7 at level 10%], while the MMC tests are significant at level 5% only for the orders 0, 1, 2, 3 [plus the orders 7, 8 and 9 at level 10%]. In view of these results and the quarterly frequency of the data, we present here results based on a VAR(4) for Granger causality testing. The results on testing Granger causality are presented in table viii.

Based on the VAR(4) model, we can identify the following significant relationships

Table vi. Power of the MMC causality tests.
VAR(1) models, $T = 30$

Φ_{ill}	$\text{MMC}_{\text{lr}}^{\text{pa}}$				
	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
0.01	3.4	1.9	3.8	2.9	1.9
0.02	3.4	4.2	7.1	6.4	4.2
0.03	7.3	7.7	23.3	15.7	6.7
0.04	10.2	15.4	41.3	30.1	14.0
0.05	18.6	22.6	65.4	49.2	25.1
0.06	26.6	37.3	78.5	65.9	41.1
0.07	40.3	54.8	88.8	77.4	53.4
0.08	50.1	63.6	95.2	87.7	63.9
0.09	62.0	77.1	97.9	94.1	75.3
0.10	70.0	84.5	98.8	96.8	82.0
0.15	91.8	98.3	100	99.9	98.8
0.20	98.9	99.8	100	100	99.7

Note – These results are based on model M0 ($\varphi = 0.90$). Power is obtained under alternatives where $\Phi_{ill} = \bar{\Phi}_{ill}(p) \neq 0$, $l = 2, \ldots, k$, $i = 1, \ldots, p$.

Table vii. Tests for VAR order (quarterly data)
AR(p) *vs.* AR($p+1$)

p	ASY_{lr}	$\text{LMC}_{\text{lr}}^{\text{pa}}$	$\text{MMC}_{\text{lr}}^{\text{pa}}$
17	0.016***	24.100	24.100
16	0.155***	33.400	33.400
15	0.312***	35.700	35.700
14	7.244*	65.100	65.100
13	12.483	70.100	70.100
12	22.961	79.000	79.000
11	2.205**	29.100	29.100
10	0.356***	11.700	11.700
9	0.058***	4.800**	6.300*
8	2.195**	4.600**	6.400*
7	4.526**	6.200*	7.900*
6	2.055**	11.300	11.300
5	8.226*	24.500	24.500
4	8.302*	20.100	20.100
3	0.916***	2.800**	4.600**
2	0.995***	2.500**	4.200**
1	0.251***	0.500***	1.100**
0	0.000***	0.100***	0.100***

Note – The numbers in the table are p-values in percentage. *** and ** highlight p-values not larger than 1% and 5%, respectively, while * highlights a p-values not larger than 10%.
– The quarterly data is computed using arithmetic average over successive blocks of three consecutive observations.

Table viii. Pairwise Granger non-causality tests based on a quarterly VAR(4) model

H_0			ASY_{lr}	$\text{LMC}_{\text{lr}}^{\text{pa}}$	$\text{MMC}_{\text{lr}}^{\text{pa}}$
M	$\nrightarrow$	r	0.495***	1.300**	2.100**
	$\nrightarrow$	P	42.195	49.600	49.600
	$\nrightarrow$	y	61.352	69.500	69.500
r	$\nrightarrow$	M	88.927	92.200	92.200
	$\nrightarrow$	P	2.108**	3.900**	5.300*
	$\nrightarrow$	y	1.671**	2.400**	3.800**
P	$\nrightarrow$	M	55.120	61.000	61.000
	$\nrightarrow$	r	22.472	29.400	29.400
	$\nrightarrow$	y	65.790	72.700	72.700
y	$\nrightarrow$	M	33.619	41.600	41.600
	$\nrightarrow$	r	0.021***	0.100***	0.200***
	$\nrightarrow$	P	25.144	33.100	33.100

Note – The numbers in the table are p-values in percentage. *** and ** highlight p-values not larger than 1% and 5%, respectively, while * highlights p-values not larger than 10%.

Table ix. Tests for VAR order (monthly data)
AR(p) *vs.* AR($p+1$)

p	ASY_{lr}	$\text{LMC}_{\text{lr}}^{\text{pa}}$	$\text{MMC}_{\text{lr}}^{\text{pa}}$
18	3.353**	16.600	16.600
17	96.756	99.200	99.200
16	7.833*	25.800	25.800
15	0.039***	0.500***	0.800***
14	0.348***	2.000**	3.300**
13	30.411	52.900	52.900
12	8.084*	20.900	20.900
11	2.279**	7.400*	8.400*
10	5.901*	13.800	13.800
9	5.549*	10.600	10.600
8	0.077***	0.300***	0.400***
7	1.984**	6.400*	7.500*
6	44.546	54.300	54.300
5	0.000***	0.100***	0.100***
4	1.938**	3.700**	4.400**
3	10.975	15.200	15.200
2	0.000***	0.100***	0.100***
1	0.000***	0.100***	0.100***
0	0.000***	0.100***	0.100***

Note – The numbers in the table are p-values in percentage. *** and ** highlight p-values not larger than 1% and 5%, respectively, while * highlights a p-values not larger than 10%.

Table x. Pairwise Granger non-causality tests based on a monthly VAR(16) model

H_0			ASY_{lr}	$\text{LMC}_{\text{lr}}^{\text{pa}}$	$\text{MMC}_{\text{lr}}^{\text{pa}}$
M	$\not\to$	r	0.004***	0.200***	0.200***
	$\not\to$	P	0.000***	0.200***	0.200***
	$\not\to$	y	2.710**	12.400	12.400
r	$\not\to$	M	14.961	33.500	33.500
	$\not\to$	P	1.997**	7.900*	8.800*
	$\not\to$	y	4.915**	16.500	16.500
P	$\not\to$	M	49.853	71.100	71.100
	$\not\to$	r	21.040	40.200	40.200
	$\not\to$	y	19.954	43.100	43.100
y	$\not\to$	M	21.543	43.700	43.700
	$\not\to$	r	0.000***	0.100***	0.100***
	$\not\to$	P	14.593	36.500	36.500

Note – The numbers in the table are p-values in percentage. *** and ** highlight p-values not larger than 1% and 5%, respectively, while * highlights a p-values not larger than 10%.

(according to MMC tests) : at level 5%,

$$M \longrightarrow r \longleftrightarrow y \tag{6.1}$$

and at level 10%

$$\begin{array}{ccccc} M & \longrightarrow & r & \longleftrightarrow & y \\ & & \downarrow & & \\ & & P & & \end{array} \tag{6.2}$$

Interestingly, these results appear to be consistent with a monetarist interpretation of the relationship between money and income, where money Granger causes interest rates which in turn causes (and is caused by) income.

For monthly data, we first studied the appropriate order [see table ix] and found that LMC and MMC tests are both significant at level 5% for the orders 0, 1, 2, 4, 5, 8, 14, 15 [plus the orders 7 and 11at level 10%]. This suggests choosing VAR(16) for analyzing the Granger causality structure between these four macroeconomic variables. The results on pairwise Granger causality tests are presented in Table x. From these results, we identify the following significant relationships (according to MMC

tests) : at level 5%,

$$\begin{array}{ccccc} M & \longrightarrow & r & \longleftarrow & y \\ & \searrow & & & \\ & & P & & \end{array} \tag{6.3}$$

and at level 10%

$$\begin{array}{ccccc} M & \longrightarrow & r & \longleftarrow & y \\ & \searrow & \downarrow & & \\ & & P & & \end{array} \tag{6.4}$$

Note that the results provided by these different models are quite similar to those of the VAR(4) model, except for the fact that we now have unidirectional (rather than bidirectional) causality from income to interest rates ($y \longrightarrow r$) and unidirectional causality from money to prices ($M \longrightarrow P$). It is interesting to note that money (when measured by bank reserves) appears not to be Granger caused by the other variables of the system. Further, money does not Granger cause income (y) directly but has an effect on the interest rate (r) which in turn Granger causes the prices. So reserves may have an indirect effect on income [see Dufour and Renault (1998) and Dufour, Pelletier, and Renault (2006)].

7. Conclusion

In this paper, we have proposed a general simulation-based method to produce finite-sample tests in parametric VAR models with known lag order (or a known upper bound on the order of the process). The method has the important feature that no other assumption is needed on the structure of the underlying process : all that is required is the possibility of simulating the model once a finite number of parameters has been specified. For example, the VAR process may be integrated of any order. We also showed that the proposed method can be implemented in practice, despite the presence of a large number of nuisance parameters. In a simulation experiment, we saw clearly that both standard asymptotic as well as bootstrap procedures can suffer from severe size distortions, while, under the same conditions, the MMC method controls

the level of the test perfectly (as expected), although its size could be lower than the test. To best of our knowledge, no other available procedure has these features. We also provided an application to Granger causality testing in a four-variable macroeconomic model of the U.S. economy.

Even though we have focused here on tests on the order of a VAR and Granger causality, the approach proposed here can be applied in principle to any set of restrictions on the model, such as unit root or cointegration hypotheses. In such cases, even though the unit root hypothesis (for example) could be taken into account by an asymptotic distributional theory or a bootstrap procedure, large roots in the stationary region but close to the unit-circle could still lead to large size distortions. By construction, the MMC procedure remains valid irrespective of the structure of the VAR. It is also important to note that the error distribution need not be normal : any assumption that specifies completely the distribution of $\varepsilon(T) = \text{vec}(\varepsilon_1, \ldots, \varepsilon_T)$, *i.e.* the disturbance distribution up to an unknown linear transformation (or covariance matrix) can be used. No assumption on the existence of moments is needed, so one could consider distributions with heavy tails. One could also introduce further free parameters in the error distribution : such parameters can be treated as extra nuisance parameters.

The main limitations of the approach proposed here lie in the parametric setup required to perform the MC tests and the computational cost. On the first issue, it is important to note that parametric assumptions involve putting a bound on the maximal order of the process (which is equivalent to assuming that the order of VAR process is "known"). In the case of testing Granger non-causality (as well as for many hypotheses of interest), this means that the lag order is an integral part of the null hypothesis : there is no way to "separate" Granger non-causality from an assumption on the order of the process. Allowing for a data-based order selection would require simulating as well the model selection procedure. Note, however, that producing finite-sample inference without putting an explicit upper bound on the order of the process is fundamentally an impossible task [see the discussions in Sims (1971a, 1971b), Cochrane (1991), Blough (1992), Faust (1996, 1999), Pötscher (2002)

and Dufour (2003)]. So, from the viewpoint of developing valid tests in finite samples, the assumption of a "known order" is unavoidable.

If one is prepared to accept a procedure which has only an "asymptotic justification", it is also important to note that the proposed "exact procedures" remain asymptotically valid (in the usual sense of pointwise asymptotic validity) under much weaker assumptions, including an "unknown" order which may be "consistently estimated". As long as the MC tests are performed using a distribution which is covered by the assumptions of the limiting distributional theory, the probability of type I error will satisfy the level condition asymptotically. Of course, under usual assumptions, such a convergence will not typically be uniform – which opens the possibility of arbitrary deviations from the nominal level of the test – but this simply reflects the fact that typical regularity assumptions are simply too weak to even allow the existence of provably valid finite-sample procedures [see Dufour (2003)]. It is worthwhile to note also that the MMC procedure automatically adapts to possible dependence of the distribution of the test statistic upon the autoregressive coefficients.

On the second issue, it is clear that MMC tests are computer intensive. The code that we used to perform the simulations and applications presented is certainly not optimal [given that these were performed with GAUSS] and we are working on improving it. Given the regular improvements in computer speeds, the importance of this type of consideration should decline in the future.

A. Appendix : Equivalence between roots and eigenvalues of a VAR process

By the definition,

$$\bar{\Phi} = \begin{bmatrix} \Phi_1 & \Phi_2 & \Phi_3 & \cdots & \Phi_{p-1} & \Phi_p \\ I_k & 0 & 0 & \cdots & 0 & 0 \\ 0 & I_k & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & I_k & 0 \end{bmatrix}, \tag{A.1}$$

we see that, for any $z \in \mathbb{C}$,

$$\det\left[I_{kp} - \bar{\Phi}z\right] = \det \begin{bmatrix} I_k - \Phi_1 z & -\Phi_2 z & -\Phi_3 z & \cdots & -\Phi_{p-1} z & -\Phi_p z \\ -I_k z & I_k & 0 & \cdots & 0 & 0 \\ 0 & -I_k z & I_k & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & I_k & 0 \\ 0 & 0 & 0 & \cdots & -I_k z & I_k \end{bmatrix}. \tag{A.2}$$

Now, on multiplying by z the second block of k columns and adding it to the first block, we get

$$\det\left[I_{kp} - \bar{\Phi}z\right] = \det \begin{bmatrix} I_k - \sum_{i=1}^{2} \Phi_i z^i & -\Phi_2 z & -\Phi_3 z & \cdots & -\Phi_{p-1} z & -\Phi_p z \\ 0 & I_k & 0 & \cdots & 0 & 0 \\ -I_k z^2 & -I_k z & I_k & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I_k & 0 \\ 0 & 0 & 0 & \cdots & -I_k z & I_k \end{bmatrix}. \tag{A.3}$$

and, repeating this process up to the p-th block (*i.e.*, multiplying the j-th block by z^{j-1} and adding the result to the first block, for $j = 2, \ldots, p$), we obtain :

$$\begin{aligned}
\det\left[I_{kp} - \bar{\Phi}z\right] &= \det\begin{bmatrix} I_k - \sum_{i=1}^{p}\Phi_i z^i & -\Phi_2 z & -\Phi_3 z & \cdots & -\Phi_{p-1}z & -\Phi_p z \\ 0 & I_k & 0 & \cdots & 0 & 0 \\ 0 & -I_k z & I_k & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I_k & 0 \\ 0 & 0 & 0 & \cdots & -I_k z & I_k \end{bmatrix} \\
&= \det\left[I_k - \sum_{i=1}^{p}\Phi_i z^i\right] = \det\left\{\Phi(z)\right\} . \qquad \text{(A.4)}
\end{aligned}$$

Since, for $z \neq 0$,

$$\det\left[I_{kp} - \bar{\Phi}z\right] = |z|^{kp}\det\left[z^{-1}I_{kp} - \bar{\Phi}\right] = (-1)^{kp}|z|^{kp}\det\left[\bar{\Phi} - z^{-1}I_{kp}\right] , \qquad \text{(A.5)}$$

the stationarity condition

$$\det\left[\Phi(z)\right] = 0 \Leftrightarrow |z| > 1 \qquad \text{(A.6)}$$

is equivalent to

$$\det\left[\bar{\Phi} - z^{-1}I_{kp}\right] = 0 \Leftrightarrow |z| > 1 \qquad \text{(A.7)}$$

and, setting $\lambda = 1/z$,

$$\det\left[\bar{\Phi} - \lambda I_{kp}\right] = 0 \Leftrightarrow |\lambda| < 1 . \qquad \text{(A.8)}$$

This means that, to have stationarity, the eigenvalues of the matrix $\bar{\Phi}$ should be *inside* the unit circle.

B. Appendix : Lag-augmented Wald tests for causality

We give here a brief description of the lag-augmented Wald tests considered in the simulation. In order to test Granger non-causality, the first step consists in estimating by ordinary least squares an unrestricted VAR($p+1$) model, rather than a VAR(p) model :

$$Y_t = \sum_{i=1}^{p+1} \Phi_i Y_{t-i} + u_t\,, \quad t = 1, \ldots, T. \tag{B.1}$$

Even though we know that $\Phi_{p+1} = 0$, this restriction is not used in order to compute the test statistic. Second, we consider the hypothesis (5.5) which is equivalent to $H_0 : (Y_2, \ldots, Y_k) \nrightarrow Y_1$ under the VAR(p) model (leaving Φ_{p+1} as a free coefficient), and compute the corresponding Wald-type test statistic $\left[\text{say } W_G^{(0)}\right]$. In accordance with (5.5), H_0 may be expressed as a set of zero restrictions on the $(p+1)\,k^2 \times 1$ coefficient vector $\phi_{p+1} = \text{vec}\left(\left[\Phi_1, \Phi_2, \ldots, \Phi_p, \Phi_{p+1}\right]'\right)$, *i.e.* $H_0' : C_{p+1}\phi_{p+1} = 0$, where C_{p+1} is a full-rank $p\,(k-1) \times (p+1)\,k^2$ matrix containing only 0 and 1. The Wald statistic then has the form :

$$W_G^{(0)} = T\big(C_{p+1}\hat{\phi}_{p+1}\big)'\left[C_{p+1}\hat{\Sigma}(\hat{\phi}_{p+1})C_{p+1}'\right]^{-1}\big(C_{p+1}\hat{\phi}_{p+1}\big) \tag{B.2}$$

where $\hat{\phi}_{p+1} = \text{vec}\left(\left[\hat{\Phi}_1, \hat{\Phi}_2, \ldots, \hat{\Phi}_p, \hat{\Phi}_{p+1}\right]'\right)$ and $\hat{\Phi}_i$, $i = 1, \ldots, p+1$, are the unconstrained least squares estimates for (B.1),[5] $\hat{\Sigma}(\hat{\phi}_{p+1})$ is the usual asymptotic covariance estimator for $T^{1/2}\big(\hat{\phi}_{p+1} - \phi_{p+1}\big)$, namely $\hat{\Sigma}(\hat{\phi}_{p+1}) = \hat{\Sigma}_{p+1} \otimes \hat{\Gamma}_{p+1}^{-1}$ with

$$\hat{\Sigma}_{p+1} = \frac{1}{T}\sum_{t=p+2}^{T} \hat{u}_t\,(p+1)\,\hat{u}_t\,(p+1)'\,, \quad \hat{\Gamma}_{p+1} = \frac{1}{T}Y\,(p+1, T)\,Y\,(p+1, T)'\,, \tag{B.3}$$

$$\hat{u}_t\,(p+1) = Y_t - \sum_{i=1}^{p+1} \hat{\Phi}_i Y_{t-i}, \quad t = p+2, \ldots, T, \tag{B.4}$$

$$Y\,(p+1, T) = \left[Y_{p+2}\,(p+1)\,, Y_{p+3}\,(p+1)\,, \ldots, Y_T\,(p+1)\right], \tag{B.5}$$

[5] Such estimates can easily be obtained by applying OLS to each equation.

and $Y_t\,(p+1) = \left[Y'_{t-1}, Y'_{t-2}, \ldots, Y'_{t-p-1}\right]'$.

Under the VAR(p) specification with H_0, this statistic follows a chi-square distribution asymptotically (with number of degrees of freedom equal to the number of restrictions) even if the process Y_t is integrated; see, for example, Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996). Of course, the finite-sample distribution of the lag-augmented Wald statistic depends on nuisance parameters (the coefficients which are not fixed by the hypotheses). So the chi-square approximation may be quite unreliable in finite samples, and improvements (such as bootstrapping) may be very important in this model. We consider here two ways of bootstrapping such lag-augmented Wald tests, a "parametric" bootstrap and a "nonparametric" one.

In the parametric case, we first obtain consistent restricted estimates $\tilde{\Phi}_i^c$, $i = 1, \ldots, p$, of the VAR(p) model [*i.e.*, (5.1) with (5.5)] along with the Cholesky factor $\tilde{R}_T^c$ associated with the estimated error covariance matrix. In the present case, the restricted estimates are obtained through maximization of the Gaussian likelihood $L(\delta)$ in (4.2). These values are then used to generate pseudo-samples $Y^{(l)} = [Y_1^{(l)}, \ldots, Y_T^{(l)}]$, according to the equation

$$Y_t^{(l)} = \sum_{i=1}^{p} \tilde{\Phi}_i^c Y_{t-i}^{(l)} + \tilde{R}^c\, \varepsilon_t^{(l)}, \quad t = 1, \ldots, T, \quad l = 1, \ldots, N, \tag{B.6}$$

where the $\varepsilon_t^{(l)}$ are simulated according to the distribution

$$\varepsilon_t^{(l)} \overset{i.i.d.}{\sim} N[0,\, I_k], \quad t = 1, \ldots, T. \tag{B.7}$$

From each simulated sample $Y^{(l)}$, a VAR($p+1$) model is estimated and the corresponding lag-augmented Wald statistic $W_G^{(l)}$ for H_0 is computed. The initial values are kept fixed at the realized values from the observed sample. The corresponding simulated p-value $\hat{p}_N\,(S_0)$ then follows according to formula (3.9) with $S_l = W_G^{(l)}$, $l = 0, 1, \ldots, N$. The null hypothesis is rejected when $\hat{p}_N\,(S_0) \leq \alpha$.

In the nonparametric case, we start from the estimated residuals

$$\tilde{u}_t^c = Y_t - \sum_{i=1}^{p} \tilde{\Phi}_i^c Y_{t-i}\,, \quad t = 1, \ldots, T. \tag{B.8}$$

New residuals $\tilde{u}_1^{(l)}, \ldots, \tilde{u}_T^{(l)}$ are then drawn at random (with replacement) from the set $\{\tilde{u}_1^c, \ldots, \tilde{u}_T^c\}$, a pseudo-sample is built following the equation

$$\tilde{Y}_t^{(l)} = \sum_{i=1}^{p} \tilde{\Phi}_i^c \tilde{Y}_{t-i}^{(l)} + \tilde{u}_t^{(l)}, \quad t = 1, \ldots, T, \tag{B.9}$$

and the corresponding lag-augmented Wald statistic for H_0 – say $\tilde{W}_G^{(l)}$ – is computed. On repeating this operation N times, the bootstrap p-value and test are finally obtained as for the parametric bootstrap.

C. Appendix : Covariance matrix coefficients used in the simulation

In section 5, the lower triangular matrices R which determine error covariance matrices $\Sigma = RR'$ were defined as follows :

$$R = \begin{bmatrix} 0.01 & 0.00 \\ -0.02 & 0.03 \end{bmatrix}, \quad \text{for } k = 2,$$

$$R = \begin{bmatrix} 0.01 & 0.00 & 0.00 \\ -0.02 & 0.03 & 0.00 \\ -0.01 & 0.01 & 0.02 \end{bmatrix}, \quad \text{for } k = 3,$$

$$R = \begin{bmatrix} 0.01 & 0.00 & 0.00 & 0.00 \\ -0.02 & 0.03 & 0.00 & 0.00 \\ -0.01 & 0.01 & 0.02 & 0.00 \\ -0.03 & 0.02 & 0.01 & 0.01 \end{bmatrix}, \quad \text{for } k = 4,$$

$$R = \begin{bmatrix} 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.02 & 0.03 & 0.00 & 0.00 & 0.00 \\ -0.01 & 0.01 & 0.02 & 0.00 & 0.00 \\ -0.03 & 0.02 & 0.01 & 0.01 & 0.00 \\ 0.01 & -0.02 & 0.03 & -0.01 & 0.02 \end{bmatrix}, \quad \text{for } k = 5,$$

$$R = \begin{bmatrix} 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.02 & 0.03 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.01 & 0.01 & 0.02 & 0.00 & 0.00 & 0.00 \\ -0.03 & 0.02 & 0.01 & 0.01 & 0.00 & 0.00 \\ 0.01 & -0.02 & 0.03 & -0.01 & 0.02 & 0.00 \\ 0.02 & -0.01 & -0.03 & 0.02 & 0.01 & 0.03 \end{bmatrix}, \quad \text{for } k = 6.$$

Chapter 2

Asymptotic distribution of a simple linear estimator for VARMA models in echelon form

1. Introduction

Multivariate time series analysis is widely based on vector autoregressive models (VAR), especially in econometric studies [see Lütkepohl (1991, 2001) and Hamilton (1994, Chapter 11)]. One reason for this popularity is that VAR models are easy to estimate and can account for relatively complex dynamic phenomena. On the other hand, very large numbers of parameters are often required to obtain a good fit, and the class of VAR models is not robust to disaggregation : if a vector process satisfies a VAR scheme, its subvectors (such as individual components) do not follow VAR processes. Instead, the subvectors of VAR processes follow vector autoregressive moving average (VARMA) processes. The latter class, indeed, includes VAR models as a special case, and can reproduce in a parsimonious way a much wider class of autocovariance structures. So they can lead to improvements in estimation and forecast precision. Further, VARMA modelling is theoretically consistent, in the sense that the subvectors of a VARMA model also satisfy VARMA schemes (usually of different order). Similarly, the VARMA class of models is not affected by temporal aggregation, while a VAR model may cease to be a VAR after it has been aggregated over time [see Lütkepohl (1987)].

VARMA modelling has been proposed a long time ago [see Hillmer and Tiao (1979), Tiao and Box (1981), Lütkepohl (1991), Boudjellaba, Dufour, and Roy (1992, 1994), Reinsel (1997)], but has remained little used in practical work. Although the process of building VARMA models is, in principle, similar to the one associated with univariate ARMA modelling, the difficulties involved are compounded by the multivariate nature of the data.

At the specification level, new identification issues (beyond the possible presence of common factors) arise and must be taken into account to ensure that unique parameter values can be associated with a given autocovariance structure (compatible with a VARMA model); see Hannan (1969b, 1970, 1976b, 1979), Deistler and Hannan (1981), Hannan and Deistler (1988, Chapter 2), Lütkepohl (1991, Chapter 7) and Reinsel (1997, Chapter 3). An important finding of this work is the importance

of the concepts of dynamic dimension and Kronecker indices in the formulation of identifiable VARMA structures. Further, specifying such models involves the selection of several autoregressive and moving average orders : in view of achieving both identifiability and efficiency, it is important that a reasonably parsimonious model be formulated. Several methods for that purpose have been proposed. The main ones include : (1) techniques based on canonical variate analysis [Akaike (1976), Cooper and Wood (1982), Tiao and Tsay (1985, 1989), Tsay (1989a)] ; (2) methods which specify an echelon form through the estimation of Kronecker indices [Hannan and Kavalieris (1984b), Tsay (1989b), Nsiri and Roy (1992, 1996), Poskitt (1992), Lütkepohl and Poskitt (1996), Bartel and Lütkepohl (1998)] ; (3) scalar-component models [Tiao and Tsay (1989), Tsay (1991)].

At the estimation level, once an identifiable specification has been formulated, the most widely proposed estimation method is maximum likelihood (ML) derived under the assumption of i.i.d. (independent and identically distributed) Gaussian innovations; see Hillmer and Tiao (1979), Tiao and Box (1981), Shea (1989), Mauricio (2002), and the review of Mélard, Roy, and Saidi (2002). This is mainly due to the presence of a moving average part in the model, which makes the latter fundamentally nonlinear. For example, in the Gaussian case, maximizing the likelihood function of a VARMA(p, q) model is typically a burdensome numerical exercise, as soon as the model includes a moving average part. Even numerical convergence may be problematic. Note also that, in the case of weak white noise innovations, quasi-maximum likelihood estimates may not be consistent. These problems also show up (at a smaller scale) in the estimation of univariate ARMA models.

From the viewpoint of making VARMA modelling, it appears crucial to have estimation methods that are both quick and simple to implement with standard statistical software, even if this may involve an efficiency cost. Another reason for putting a premium on such estimation methods is that large-sample distributional theory tends to be quite unreliable in high-dimensional dynamic models, so that tests and confidence sets based on asymptotic approximations are also unreliable (for example, the actual size of test procedures may be far larger than their nominal size).

This suggests that simulation-based procedures – for example, bootstrap techniques – should be used, but simulation may be impractical if calculation of the estimators involved is difficult or time consuming.

In the case of univariate ARMA models, a relatively simple estimation procedure was originally proposed by Hannan and Rissanen (1982); see also Durbin (1960), Hannan and Kavalieris (1984a), Zhao-Guo (1985), Hannan, Kavalieris, and Mackisack (1986), Poskitt (1987), Koreisha and Pukkila (1990a, 1990b, 1995), Pukkila, Koreisha, and Kallinen (1990) and Galbraith and Zinde-Walsh (1994, 1997). This approach is based on estimating (by least squares) the innovations of the process through a long autoregression; after that, the lagged innovations are replaced by the corresponding residuals in the ARMA equation, which may then be also estimated by least squares.

Extensions of this method to VARMA models have been studied by Hannan and Kavalieris (1984b, 1986), Hannan and Deistler (1988), Koreisha and Pukkila (1989), Huang and Guo (1990), Poskitt (1992), Poskitt and Lütkepohl (1995), Lütkepohl and Poskitt (1996), Lütkepohl and Claessen (1997) and Flores de Frutos and Serrano (2002). Work on VARMA estimation has focused on preliminary use of such linear estimators for model selection purposes. It is then suggested that other estimation procedures (such as ML) be used. Although consistency is proved, the asymptotic distribution of the basic two-step estimator has not apparently been supplied.

In this paper, we consider the problem of estimating the parameters of stationary invertible VARMA models in echelon form using only linear least squares methods. The echelon form is selected because it tends to deliver relatively parsimonious parameterizations. In particular, we study a simple two-step estimator that can be implemented only through single equation linear regressions and thus is remarkably simple to apply. Such an estimator was previously considered in the above mentioned work on linear VARMA estimation, but its asymptotic distribution has not apparently been established. Given the Kronecker indices of the VARMA process, we derive the asymptotic distribution of this estimator under standard regularity conditions. In particular, we show that the latter has an asymptotic normal distribution (which entails its consistency), and we provide a simple consistent estimator for its asymptotic

covariance matrix, so that asymptotically valid tests and confidence tests can be built for the parameters of the model.

The paper is organized as follows. In section **2**, we formulate the background model, where the echelon form VARMA representation is considered to ensure unique parametrization, and we define the assumptions which will be used in the rest of the paper. The two-step linear estimation procedure studied in the paper is described in section **3**, and we derive its asymptotic distribution in section **4**. We conclude in section **5**. The proofs of the propositions and theorems appear in the Appendix.

2. Framework

In this section, we describe the theoretical framework and the assumptions we will consider in the sequel. We will first define the standard VARMA representation. As the latter may involve identification problems, we will then define the echelon form on the VARMA model, which ensures uniqueness of model parameters. Finally, we shall formulate the basic regularity assumptions we shall consider.

2.1. Standard form

A k-dimensional regular vector process $\{Y_t : t \in \mathbb{Z}\}$ has a VARMA(p, q) representation if it satisfies an equation of the form :

$$Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + u_t + \sum_{j=1}^{q} B_j u_{t-j}, \tag{2.1}$$

for all t, where $Y_t = (Y_{1,t}, \ldots, Y_{k,t})'$, p and q are non-negative integers (respectively, the autoregressive and moving average orders), A_i and B_j the $k \times k$ coefficient matrices, and $\{u_t : t \in \mathbb{Z}\}$ is a (second order) white noise $WN[0, \Sigma_u]$, where Σ_u is a $k \times k$ positive definite symmetric matrix. Under the stationarity and invertibility conditions the coefficients A_i and B_j satisfy the constraints

$$\det\{A(z)\} \neq 0 \text{ and } \det\{B(z)\} \neq 0 \text{ for all } |z| \leq 1 \tag{2.2}$$

where z is a complex number, $A(z) = I_k - \sum_{i=1}^{p} A_i z^i$ and $B(z) = I_k + \sum_{j=1}^{q} B_j z^j$. This process has the following autoregressive and moving average representations :

$$Y_t = \sum_{\tau=1}^{\infty} \Pi_\tau Y_{t-\tau} + u_t, \tag{2.3}$$

$$Y_t = u_t + \sum_{v=1}^{\infty} \Psi_v u_{t-v}, \ t = 1, \ldots, T, \tag{2.4}$$

where

$$\Pi(z) = B(z)^{-1} A(z) = I_k - \sum_{\tau=1}^{\infty} \Pi_\tau z^\tau, \tag{2.5}$$

$$\Psi(z) = A(z)^{-1} B(z) = I_k + \sum_{v=1}^{\infty} \Psi_v z^v, \tag{2.6}$$

$$\det\{\Pi(z)\} \neq 0 \text{ and } \det\{\Psi(z)\} \neq 0, \text{ for all } |z| \leq 1. \tag{2.7}$$

Note also that we can find real constants $C > 0$ and $\rho \in (0, 1)$ such that

$$\|\Pi_\tau\| \leq C\rho^\tau \text{ and } \|\Psi_v\| \leq C\rho^v, \tag{2.8}$$

hence

$$\sum_{\tau=1}^{\infty} \|\Pi_\tau\| < \infty, \quad \sum_{v=1}^{\infty} \|\Psi_v\| < \infty, \tag{2.9}$$

where $\|.\|$ is the Schur norm for a matrix [see Horn and Johnson (1985, section 5.6)], *i.e.*

$$\|M\|^2 = \text{tr}[M'M]. \tag{2.10}$$

2.2. Echelon form

It is well known that the standard VARMA(p, q) representation given by (2.1) is not unique, in the sense that different sets of coefficients A_i and B_j may represent the same autocovariance structure. To ensure a unique parameterization, we shall consider the stationary invertible VARMA(p, q) process in echelon form representation. Such

a representation can be defined as follows :

$$\Phi(L) Y_t = \Theta(L) u_t, \tag{2.11}$$

$$\Phi(L) = \Phi_0 - \sum_{i=1}^{\bar{p}} \Phi_i L^i, \quad \Theta(L) = \Theta_0 + \sum_{j=1}^{\bar{p}} \Theta_j L^j, \tag{2.12}$$

where L denotes the lag operator, $\Phi_i = \left[\phi_{lm,i}\right]_{l,m=1,\ldots,k}$ and $\Theta_j = \left[\theta_{lm,j}\right]_{l,m=1,\ldots,k}$, $\bar{p} = \max(p, q)$, $\Theta_0 = \Phi_0$, and Φ_0 is a lower-triangular matrix whose diagonal elements are all equal to one. The VARMA representation (2.11) has an echelon form if $\Phi(L) = \left[\phi_{lm}(L)\right]_{l,m=1,\ldots,k}$ and $\Theta(L) = \left[\theta_{lm}(L)\right]_{l,m=1,\ldots,k}$ satisfy the following conditions : given a vector of orders $(p_1, \ldots, p_k)$ called the *Kronecker indices*, the operators $\phi_{lm}(L)$ and $\theta_{lm}(L)$ on any given row l of $\Phi(L)$ and $\Theta(L)$ have the same degree p_l $(1 \leq l \leq k)$ and

$$\begin{aligned} \phi_{lm}(L) &= 1 - \sum_{i=1}^{p_l} \phi_{ll,i} L^i \qquad \text{if } l = m, \\ &= - \sum_{i=p_l - p_{lm}+1}^{p_l} \phi_{lm,i} L^i \quad \text{if } l \neq m, \end{aligned} \tag{2.13}$$

$$\theta_{lm}(L) = \sum_{j=0}^{p_l} \theta_{lm,j} L^j \quad \text{with } \Theta_0 = \Phi_0, \tag{2.14}$$

for $l, m = 1, \ldots, k$, where

$$\begin{aligned} p_{lm} &= \min(p_l + 1, p_m) \quad \text{for } l \geq m, \\ &= \min(p_l, p_m) \qquad \text{for } l < m. \end{aligned} \tag{2.15}$$

Clearly, $p_{ll} = p_l$ is the order of the polynomial (*i.e.*, the number of free coefficients) on the l-th diagonal element of $\Phi(L)$ as well as the order of the polynomials on the corresponding row of $\Theta(L)$, while p_{lm} specifies the number of free coefficients in the operator $\phi_{lm}(L)$ for $l \neq m$. The sum of the Kronecker indices $\sum_{l=1}^{k} p_l$ is called the McMillan degree. The P matrix formed by the Kronecker indices associated with the model is $P = \left[p_{lm}\right]_{l,m=1,\ldots,k}$. This leads to $\sum_{l=1}^{k} \sum_{m=1}^{k} p_{lm}$ autoregressive and $k \sum_{l=1}^{k} p_l$ moving average free coefficients, respectively. Obviously, for the VARMA orders we have $\bar{p} = \max(p_1, \ldots, p_k)$. Note that this identified parameterization for VARMA(p, q)

models ensures the uniqueness of left-coprime operators $\Phi(L)$ and $\Theta(L)$. Although other identifiable parameterizations could be used – such as the final equations form – the echelon form tends to be more parsimonious and can lead to efficiency gains. For proofs of the uniqueness of the echelon form and for other identification conditions, the reader should consult to Hannan (1969b, 1970, 1976a, 1979), Deistler and Hannan (1981), Hannan and Deistler (1988) and Lütkepohl (1991, Chapter 7).

The stationarity and invertibility conditions for echelon form of (2.11) are the same as usual, namely

$$\det\{\Phi(z)\} \neq 0 \quad \text{for all } |z| \leq 1, \tag{2.16}$$

for stationarity, and

$$\det\{\Theta(z)\} \neq 0 \quad \text{for all } |z| \leq 1, \tag{2.17}$$

for invertibility, where

$$\Phi(z) = \Phi_0 - \sum_{i=1}^{\bar{p}} \Phi_i z^i, \quad \Theta(z) = \Theta_0 + \sum_{j=1}^{\bar{p}} \Theta_j z^j, \tag{2.18}$$

with $\Pi(z) = \Theta(z)^{-1}\Phi(z)$ and $\Psi(z) = \Phi(z)^{-1}\Theta(z)$. It will be useful to observe that (2.11) can be rewritten in the following form :

$$Y_t = (I_k - \Phi_0) V_t + \sum_{i=1}^{\bar{p}} \Phi_i Y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j u_{t-j} + u_t \tag{2.19}$$

where

$$V_t = Y_t - u_t = \Phi_0^{-1}\Big[\sum_{i=1}^{\bar{p}} \Phi_i Y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j u_{t-j}\Big]. \tag{2.20}$$

Note that V_t is a function of *lagged* values of Y_t and u_t, so that the error term u_t in (2.19) is uncorrelated with all the other variables on the right-hand side of the equation.

Set

$$X_t = \left[V_t', Y_{t-1}', \ldots, Y_{t-\bar{p}}', u_{t-1}', \ldots, u_{t-\bar{p}}'\right]', \tag{2.21}$$

$$D = \left[I_k - \Phi_0, \Phi_1, \ldots, \Phi_{\bar{p}}, \Theta_1, \ldots, \Theta_{\bar{p}}\right]'. \tag{2.22}$$

The vector X_t has dimension $(kh) \times 1$ where $h = 2\bar{p}+1$ while D is a $(kh) \times k$ matrix of coefficients. In view of (2.20), it is clear the covariance matrix of X_t is singular, so it is crucial that (identifying) restrictions be imposed on model coefficients. Under the restrictions of the echelon form (2.12) - (2.15), we can find a unique $(k^2h) \times \nu$ full rank matrix R such that $\beta = R\eta$, where η is a $\nu \times 1$ vector of free coefficients and $\nu < k^2h$. Thus Y_t in (2.19) can be expressed as

$$Y_t = D'X_t + u_t = [I_k \otimes X_t'] R\eta + u_t. \tag{2.23}$$

The structure of R is such that

$$\beta = \mathrm{vec}[D] = R\eta, \tag{2.24}$$

$$R = \mathrm{diag}(R_1, \ldots, R_k) = \begin{bmatrix} R_1 & 0 & \cdots & 0 \\ 0 & R_2 & \cdots & \vdots \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & R_k \end{bmatrix}, \tag{2.25}$$

where R_i, $i = 1, 2, \ldots, k$, are $(kh) \times \nu_i$ full-rank selection (zero-one) matrices, each one of which selects the non-zero elements of the corresponding equation, and ν_i is the number of freely varying coefficients present in the i-th equation. The structure of R_i is such that $R_i'R_i = I_{\nu_i}$ and $\beta_i = R_i\eta_i$ where β_i and η_i are respectively a $(kh) \times 1$ and $\nu_i \times 1$ vectors so that β_i is the unconstrained parameter vector in the i-th equation of (2.19) – on which zero restrictions are imposed – and η_i is the corresponding vector of free parameters :

$$\beta = \left(\beta_1', \beta_2', \ldots, \beta_k'\right)', \quad \eta = \left(\eta_1', \eta_2', \ldots, \eta_k'\right)'. \tag{2.26}$$

Note also that successful identification entails that

$$\operatorname{rank}\{\mathsf{E}(R'[I_k \otimes X_t][I_k \otimes X_t']R)\} = \operatorname{rank}\{R'[I_k \otimes \Gamma]R\} = \nu \tag{2.27}$$

where $\Gamma = \mathsf{E}[X_t X_t']$, or equivalently

$$\operatorname{rank}\{\mathsf{E}[R_i' X_t X_t' R_i]\} = \operatorname{rank}\{R_i' \Gamma R_i\} = \nu_i, \quad i = 1, \ldots, k. \tag{2.28}$$

Setting

$$\begin{aligned} X(T) &= [X_1, \ldots, X_T]', && (2.29)\\ Y(T) &= [Y_1, \ldots, Y_T]' = [y_1(T), \ldots, y_k(T)], && (2.30)\\ U(T) &= [u_1, \ldots, u_T]' = [U_1(T), \ldots, U_k(T)], && (2.31)\\ y(T) &= \operatorname{vec}[Y(T)], \quad u(T) = \operatorname{vec}[U(T)], && (2.32) \end{aligned}$$

(2.23) can be put in any one of the two following matrix forms :

$$\begin{aligned} Y(T) &= X(T)D + U(T), && (2.33)\\ y(T) &= [I_k \otimes X(T)]R\eta + u(T), && (2.34) \end{aligned}$$

where $[I_k \otimes X(T)]R$ is a $(kT) \times \nu$ matrix. In the sequel, we shall assume that

$$\operatorname{rank}\{[I_k \otimes X(T)]R\} = \nu \text{ with probability 1.} \tag{2.35}$$

Under the assumption that the process is a regular process with continuous distribution, it is easy that the latter must hold.

To see better how the echelon restrictions should be written, consider the following VARMA$(2,1)$ model in echelon form :

$$\begin{aligned} Y_{1,t} &= \phi_{11,1} Y_{1,t-1} + \phi_{11,2} Y_{1,t-2} + u_{1,t}, && (2.36)\\ Y_{2,t} &= \phi_{21,0}(Y_{1,t} - u_{1,t}) + \phi_{21,1} Y_{1,t-1} + \phi_{22,1} Y_{2,t-1} + \theta_{22,1} u_{2,t-1} + u_{2,t}. && (2.37) \end{aligned}$$

In this case, we have :

$$\Phi(L) = \begin{bmatrix} 1-\phi_{11,1}L-\phi_{11.2}L^2 & -\phi_{12,2}L^2 \\ -\phi_{21,0}-\phi_{21,1}L & 1-\phi_{22,1}L \end{bmatrix}, \tag{2.38}$$

$$\Theta(L) = \begin{bmatrix} 1+\theta_{11,1}L+\theta_{11,2}L^2 & \theta_{12,1}L+\theta_{12,2}L^2 \\ \theta_{21,1}L & 1+\theta_{22,1}L \end{bmatrix}, \tag{2.39}$$

with $\phi_{12,2}=0$, $\theta_{11,1}=0$, $\theta_{11,2}=0$, $\theta_{12,1}=0$, $\theta_{12,2}=0$, $\theta_{21,1}=0$, so that the Kronecker indices are $p_1=p_{11}=2$, $p_2=p_{22}=1$, $p_{21}=2$ and $p_{12}=1$. Setting $X_t=\left[V_t',Y_{t-1}',Y_{t-2}',u_{t-1}'\right]'$, $V_t=(V_{1,t},V_{2,t})'$, $V_{1,t}=(Y_{1,t}-u_{1,t})$ and $V_{2,t}=(Y_{2,t}-u_{2,t})$, we can then write :

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \phi_{21,0} & 0 \end{bmatrix}\begin{bmatrix} V_{1,t} \\ V_{2,t} \end{bmatrix} + \begin{bmatrix} \phi_{11,1} & 0 \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix}\begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \phi_{11,2} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} Y_{1,t-2} \\ Y_{2,t-2} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \theta_{22,1} \end{bmatrix}\begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}. \tag{2.40}$$

Here we have :

$$\beta = \left(0,0,\phi_{11,1},0,\phi_{11,2},0,0,0,\phi_{21,0},0,\phi_{21,1},\phi_{22,1},0,0,0,\theta_{22,1}\right)', \tag{2.41}$$

$$\eta = \left(\phi_{11,1},\phi_{11,2},\phi_{21,0},\phi_{21,1},\phi_{22,1},\theta_{22,1}\right)', \tag{2.42}$$

$$\left[I_k \otimes X_t'\right] R = \begin{bmatrix} Y_{1,t-1} & Y_{1,t-2} & 0 & 0 & 0 & 0 \\ 0 & 0 & V_{1,t} & Y_{1,t-1} & Y_{2,t-1} & u_{2,t-1} \end{bmatrix}, \tag{2.43}$$

and

$$\big[I_k \otimes X(T)\big]R = \begin{bmatrix} Y_{1,0} & Y_{1,-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & V_{1,1} & Y_{1,0} & Y_{2,0} & u_{2,0} \\ Y_{1,1} & Y_{1,0} & 0 & 0 & 0 & 0 \\ 0 & 0 & V_{1,2} & Y_{1,1} & Y_{2,1} & u_{2,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{1,T-1} & Y_{1,T-2} & 0 & 0 & 0 & 0 \\ 0 & 0 & V_{1,T} & Y_{1,T-1} & Y_{2,T-1} & u_{2,T-1} \end{bmatrix}. \tag{2.44}$$

The appropriate matrix R is given by :

$$R' = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.45}$$

2.3. Regularity assumptions

In order to establish the asymptotic distribution of the linear estimator defined below, we will need further assumptions on the innovation process and the truncation lag of the first step autoregression. We now state the assumptions we shall consider.

Assumption 2.1 STRONG WHITE NOISE INNOVATIONS. *The vectors u_t, $t \in \mathbb{Z}$, are independent and identically distributed (i.i.d.) with mean zero, covariance matrix Σ_u and continuous distribution.*

Assumption 2.2 UNIFORM BOUNDEDNESS OF FOURTH MOMENTS. *There is a finite constant m_4 such that, for all $1 \leq i,\ j,\ r,\ s \leq k$ and for all t,*

$$\mathsf{E}\,|u_{it}u_{jt}u_{rt}u_{st}| \leq m_4 < \infty.$$

Assumption 2.3 AUTOREGRESSIVE TRUNCATION LAG OF ORDER LESS THAN $T^{1/2}$. *n_T is a function of T such that*

$$n_T \to \infty \text{ and } n_T^2/T \to 0 \text{ as } T \to \infty \tag{2.46}$$

and, for some $c > 0$ and $0 < \bar{\delta} < 1/2$,

$$n_T \geq cT^{\bar{\delta}} \text{ for } T \text{ sufficiently large.} \tag{2.47}$$

Assumption 2.4 DECAY RATE OF TRUNCATED AUTOREGRESSIVE COEFFICIENTS. *The coefficients of the autoregressive representation* (2.3)

$$n_T^{1/2} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T \to \infty. \tag{2.48}$$

Assumption **2.1** means that we have a strong VARMA process, while Assumption **2.2** on moments of order four will ensure the empirical autocovariances of the process have finite variances. Assumption **2.3** implies that n_T goes to infinity at a rate slower than $T^{1/2}$; for example, the assumption is satisfied if $n_T = cT^{\delta}$ with $0 < \bar{\delta} \leq \delta < 1/2$. Assumption **2.4** characterizes the rate of decay of autoregressive coefficients in relation with n_T.

Although the above assumptions are sufficient to show consistency of the two-stage linear estimator, we will need another assumption to show that the asymptotic distribution is normal with a distribution which is unaffected by the use of estimated innovations.

Assumption 2.5 AUTOREGRESSIVE TRUNCATION LAG OF ORDER LESS THAN $T^{1/4}$. *n_T is a function of T such that*

$$n_T \to \infty \text{ and } n_T^4/T \to 0 \text{ as } T \to \infty. \tag{2.49}$$

The latter assumption means that n_T goes to infinity at a rate slower than $T^{1/4}$; for example, it is satisfied if $n_T = cT^{\delta}$ with $0 < \bar{\delta} \leq \delta < 1/4$. It is easy to see that the condition (2.49) entails (2.46). Finally, it is worthwhile to note that (2.48) holds for VARMA processes whenever $n_T = cT^{\delta}$ with $c > 0$ and $\delta > 0$, *i.e.*

$$T^{\delta} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T \to \infty, \quad \text{for all } \delta > 0. \tag{2.50}$$

This is easy to see from the exponential decay property of VARMA processes [see (2.8)].

3. Two-step linear estimation

In this section, we describe a simple estimation procedure for a VARMA models in echelon form with known order. The Kronecker indices characterizing the echelon form VARMA model are taken as given, and we focus our attention on the estimation of the autoregressive and moving average coefficients.

Let $(Y_{-n_T+1}, \ldots, Y_T)$ be a random sample of size $T + n_T$, where n_T goes to infinity as T goes to infinity. We consider first a "long" multivariate linear vector autoregression :

$$Y_t = \sum_{\tau=1}^{n_T} \Pi_\tau Y_{t-\tau} + u_t(n_T), \quad t = 1, \ldots, T, \tag{3.1}$$

and the corresponding least squares estimates :

$$\tilde{\Pi}(n_T) = \left[\tilde{\Pi}_1(n_T), \ldots, \tilde{\Pi}_{n_T}(n_T)\right]. \tag{3.2}$$

Such an estimation can be performed by running k separate univariate linear regressions (one for each variable in Y_t). Yule-Walker estimates of the corresponding theoretical coefficients Π_τ could also be considered. Then, under model (2.3) and the assumptions **2.1** to **2.4**, it follows from the results of Paparoditis (1996, Theorem 2.1)

and Lewis and Reinsel (1985, proof of Theorem 1) that :

$$\left\| \tilde{\Pi}\left(n_T\right) - \Pi\left(n_T\right) \right\| = O_p\left(n_T^{1/2}/T^{1/2}\right) \tag{3.3}$$

where

$$\Pi\left(n_T\right) = \left[\Pi_1, \ldots, \Pi_{n_T}\right]. \tag{3.4}$$

As usual, for any sequence of random variables Z_T and positive numbers r_T, $T = 1, 2, \ldots,$ the notation $Z_T = O_p(r_T)$ means that Z_T/r_T is asymptotically bounded in probability (as $T \to \infty$), while $Z_T = o_p(r_T)$ means that Z_T/r_T converges to zero in probability. When Y_t satisfies a VARMA scheme, the assumptions **2.3** and **2.4** are satisfied by any truncation lag of the form $n_T = cT^{\delta}$ with $c > 0$ and $0 < \delta < 1/2$. If, furthermore, the assumptions **2.3** and **2.4** are replaced by stronger ones, namely

$$n_T \to \infty \text{ and } n_T^3/T \to 0 \text{ as } T \to \infty, \tag{3.5}$$

$$T^{1/2} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T \to \infty, \tag{3.6}$$

then asymptotic normality also holds :

$$T^{1/2}\, l\left(n_T\right)' \left[\tilde{\pi}\left(n_T\right) - \pi\left(n_T\right)\right] \underset{T\to\infty}{\longrightarrow} \mathrm{N}\left[0,\, l\left(n_T\right)' Q(n_T) l\left(n_T\right)\right], \tag{3.7}$$

where $l\left(n_T\right)$ is a sequence of $k^2 n_T \times 1$ vectors such that $0 < M_1 \leq \|l\left(n_T\right)\| \leq M_2 < \infty$ for $n_T = 1, 2, \ldots,$ and

$$\tilde{\pi}\left(n_T\right) - \pi\left(n_T\right) = \mathrm{vec}\left[\tilde{\Pi}\left(n_T\right) - \Pi\left(n_T\right)\right], \tag{3.8}$$

$$Q(n_T) = \Gamma(n_T)^{-1} \otimes \Sigma_u,\ \Gamma(n_T) = \mathsf{E}\left[Y_t(n_T) Y_t(n_T)'\right], \tag{3.9}$$

$$Y_t(n_T) = \left[Y_{t-1}', Y_{t-2}', \ldots, Y_{t-n_T}'\right]'. \tag{3.10}$$

Note that a possible choice for the sequence n_T that satisfies both $n_T^3/T \to 0$ and $T^{1/2} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0$ is for example $n_T = T^{1/\varepsilon}$ with $\varepsilon > 3$. On the other hand $n_T = \ln(\ln T)$, as suggested by Hannan and Kavalieris (1984b), is not a permissible

choice because in general $T^{1/2}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|$ does not approach zero as $T\to\infty$.

Let

$$\tilde{u}_t(n_T) = Y_t - \sum_{\tau=1}^{n_T}\tilde{\Pi}_\tau(n_T)Y_{t-\tau} = Y_t - \tilde{\Pi}\left(n_T\right)Y_t(n_T) \tag{3.11}$$

be the estimated residuals obtained from the first stage estimation procedure,

$$\tilde{\Sigma}_u(n_T) = \frac{1}{T}\sum_{t=1}^{T}\tilde{u}_t(n_T)\tilde{u}_t(n_T)' \tag{3.12}$$

the corresponding estimator of the innovation covariance matrix, and

$$\hat{\Sigma}_T = \frac{1}{T}\sum_{t=1}^{T}u_t u_t' \tag{3.13}$$

the covariance "estimator" based on the true innovations. Then, we have the following equivalences and convergences.

Proposition 3.1 INNOVATION COVARIANCE ESTIMATOR CONSISTENCY. *Let* $\{Y_t : t\in\mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VARMA echelon representation given by* (2.11) - (2.15)*. Then, under the assumptions* **2.1** *to* **2.4**, *we have :*

$$\left\|\frac{1}{T}\sum_{t=1}^{T}u_t\left[\tilde{u}_t(n_T)-u_t\right]'\right\| = O_p\left(\frac{n_T}{T}\right), \tag{3.14}$$

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{u}_t(n_T)-u_t\right\|^2 = O_p\left(\frac{n_T^2}{T}\right), \tag{3.15}$$

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{u}_t(n_T)-u_t\right]\left[\tilde{u}_t(n_T)-u_t\right]'\right\| = O_p\left(\frac{n_T^2}{T}\right), \tag{3.16}$$

$$\left\|\tilde{\Sigma}_u(n_T)-\hat{\Sigma}_T\right\| = O_p\left(\frac{n_T^2}{T}\right), \quad \left\|\tilde{\Sigma}_u(n_T)-\Sigma_u\right\| = O_p\left(\frac{n_T^2}{T}\right). \tag{3.17}$$

The asymptotic equivalence between $\tilde{u}_t(n_T)$ and u_t stated in the above proposition suggests we may be able to consistently estimate the parameters of the VARMA model in (2.19) after replacing the unobserved lagged innovations $u_{t-1},\ldots,u_{t-\bar{p}}$ with the corresponding residuals $\tilde{u}_{t-1}(n_T),\ldots,\tilde{u}_{t-\bar{p}}(n_T)$ from the above long autoregression.

So, in order to estimate the coefficients Φ_i and Θ_j of the VARMA process, we consider a linear regression of the form

$$Y_t = \sum_{i=1}^{\bar{p}} \Phi_i Y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j \tilde{u}_{t-j}(n_T) + e_t(n_T) \tag{3.18}$$

imposing the (exclusion) restrictions associated with the echelon form. Setting

$$\tilde{V}_t(n_T) = Y_t - \tilde{u}_t(n_T), \tag{3.19}$$

this regression can also be put in a regression form similar to (2.19) :

$$Y_t = (I_k - \Phi_0)\, \tilde{V}_t(n_T) + \sum_{i=1}^{\bar{p}} \Phi_i Y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j \tilde{u}_{t-j}(n_T) + e_t(n_T) \tag{3.20}$$

where

$$e_t(n_T) = \tilde{u}_t(n_T) + \sum_{j=0}^{\bar{p}} \Theta_j \big[u_{t-j} - \tilde{u}_{t-j}(n_T)\big]. \tag{3.21}$$

Note that (3.20) can be written as

$$Y_t = \big[I_k \otimes \tilde{X}_t(n_T)'\big] R\eta + e_t(n_T), \quad t = 1, \ldots, T, \tag{3.22}$$

where

$$\tilde{X}_t(n_T) = \big[\tilde{V}_t(n_T)', Y'_{t-1}, \ldots, Y'_{t-\bar{p}}, \tilde{u}_{t-1}(n_T)', \ldots, \tilde{u}_{t-\bar{p}}(n_T)'\big]'. \tag{3.23}$$

Therefore, the second step estimators $\tilde{\eta}$ can be obtained by running least squares on the equations (3.22). Setting

$$\tilde{X}(n_T) = \big[\tilde{X}_1(n_T), \tilde{X}_2(n_T), \ldots, \tilde{X}_T(n_T)\big]' \tag{3.24}$$

we get, after some manipulations,

$$\begin{aligned} \tilde{\eta} &= \Big\{R'\big[I_k \otimes \tilde{X}(n_T)'\tilde{X}(n_T)\big]R\Big\}^{-1} R'\big[I_k \otimes \tilde{X}(n_T)'\big] y(T) \\ &= \big(\tilde{\eta}'_1, \tilde{\eta}'_2, \ldots, \tilde{\eta}'_k\big)' \end{aligned} \tag{3.25}$$

where

$$\tilde{\eta}_i = \left\{R_i'\tilde{X}(n_T)'\tilde{X}(n_T)R_i\right\}^{-1}R_i'\tilde{X}(n_T)'y_i(T). \tag{3.26}$$

$\tilde{\eta}$ can be easily obtained by stacking the single equation LS estimators $\tilde{\eta}_i$ which are obtained by regressing y_i on $\tilde{X}(n_T)R_i$.

4. Asymptotic distribution

We will now study the asymptotic distribution of the linear estimator described in the previous section. For that purpose, we note first that the estimator $\tilde{\eta}$ in (3.25) can be expressed as

$$\tilde{\eta} = \left\{R'\left[I_k \otimes \tilde{\Gamma}(n_T)\right]R\right\}^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T} R'\left[I_k \otimes \tilde{X}_t(n_T)\right]Y_t\right\} \tag{4.1}$$

where

$$\tilde{\Gamma}(n_T) = \frac{1}{T}\sum_{t=1}^{T}\tilde{X}_t(n_T)\tilde{X}_t(n_T)'. \tag{4.2}$$

Let also

$$\tilde{\Upsilon}(n_T) = I_k \otimes \tilde{\Gamma}(n_T), \quad \tilde{Q}(n_T) = \left\{R'\tilde{\Upsilon}(n_T)R\right\}^{-1}, \tag{4.3}$$

$$\tilde{\Omega}(n_T) = \frac{1}{T}\sum_{t=1}^{T} R'\left[I_k \otimes \tilde{X}_t(n_T)\right]e_t(n_T). \tag{4.4}$$

It is then easy to see that

$$\tilde{\eta} - \eta = \tilde{Q}(n_T)\tilde{\Omega}(n_T) \tag{4.5}$$

hence

$$\left\|\tilde{\eta} - \eta\right\| \leq \left\|\tilde{Q}(n_T)\right\|_1\left\|\tilde{\Omega}(n_T)\right\| \leq \left\|\tilde{Q}(n_T)\right\|\left\|\tilde{\Omega}(n_T)\right\| \tag{4.6}$$

where $\|A\|_1 = \sup_{x\neq 0}\left\{\frac{\|Ax\|}{\|x\|}\right\}$ stands for the largest eigenvalue of $A'A$ and we used the inequality $\|AB\|^2 \leq \|A\|_1^2\|B\|^2$ for any two conformable matrices A and B [see Horn and Johnson (1985, section 5.6)].

Define

$$\Upsilon = I_k \otimes \Gamma, \quad Q = \left\{R'\Upsilon R\right\}^{-1}, \tag{4.7}$$

$$\Gamma_T = \frac{1}{T}\sum_{t=1}^{T} X_t X_t', \quad \Upsilon_T = I_k \otimes \Gamma_T = \frac{1}{T}\sum_{t=1}^{T}\left[I_k \otimes X_t X_t'\right], \tag{4.8}$$

$$Q_T = \left\{R'\Upsilon_T R\right\}^{-1}, \quad \Omega_T = \frac{1}{T}\sum_{t=1}^{T} R'\left[I_k \otimes X_t\right] u_t. \tag{4.9}$$

Note that $R'\Upsilon R$ is positive definite by the regularity assumption. To study the convergence and distributional properties of $\tilde{\eta} - \eta$, we need first to establish the following proposition.

Proposition 4.1 GENERAL CONDITION FOR BOUNDED SYMMETRIC MATRIX 2. *Let $\{Y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA echelon representation given by* (2.11) - (2.15). *Then, under the assumptions* **2.1** *to* **2.4**, *we have the following equivalences :*

$$\frac{1}{T}\left\|\tilde{X}(n_T) - X(T)\right\|^2 = O_p\left(\frac{n_T^2}{T}\right), \tag{4.10}$$

$$\left\|\tilde{\Gamma}(n_T) - \Gamma_T\right\| = O_p\left(\frac{n_T}{T^{1/2}}\right), \tag{4.11}$$

$$\left\|\tilde{\Upsilon}(n_T) - \Upsilon_T\right\| = O_p\left(\frac{n_T}{T^{1/2}}\right), \tag{4.12}$$

$$\left\|\tilde{Q}(n_T)^{-1} - Q^{-1}\right\| = O_p\left(\frac{n_T}{T^{1/2}}\right), \tag{4.13}$$

$$\left\|\tilde{Q}(n_T) - Q\right\| = O_p\left(\frac{n_T}{T^{1/2}}\right). \tag{4.14}$$

The latter proposition shows that the matrices $\tilde{\Gamma}(n_T)$, $\tilde{\Upsilon}(n_T)$, $\tilde{Q}(n_T)^{-1}$ and $\tilde{Q}(n_T)$ – based on approximate innovations (estimated from a long autoregression) – are all asymptotically equivalent to the corresponding matrices based on true innovations, according to the rate $n_T/T^{1/2}$. Similarly, the norm of the difference between the approximate regressor matrix $\tilde{X}(n_T)$ and $X(T)$ has order $O_p(n_T/T^{1/2})$. This suggests that $\tilde{\eta}$ converges to η, and we give the appropriate rate of convergence in the following

theorem.

Theorem 4.1 CONSISTENCY OF SECOND STEP HR ESTIMATES. *Let* $\{Y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VARMA echelon representation given by* (2.11) - (2.15). *Then, under the assumptions* **2.1** *to* **2.4**, *we have*

$$\|\Omega_T\| = O_p\left(\frac{1}{T^{1/2}}\right), \quad \|\tilde{\Omega}(n_T) - \Omega_T\| = O_p\left(\frac{n_T^2}{T}\right), \tag{4.15}$$

$$\|\tilde{\eta} - \eta\| = O_p\left(\frac{1}{T^{1/2}}\right) + O_p\left(\frac{n_T^2}{T}\right). \tag{4.16}$$

If, furthermore,

$$n_T^4/T \to 0 \text{ as } T \to \infty, \tag{4.17}$$

then

$$\|\tilde{\eta} - \eta\| = O_p\left(\frac{1}{T^{1/2}}\right). \tag{4.18}$$

The latter theorem shows that $\tilde{\eta}$ is a consistent estimator. If furthermore, $n_T^4/T \to 0$ as $T \to \infty$, then $\tilde{\eta}$ converges at the rate $T^{-1/2}$ which is typically expected to get asymptotic normality. In order to derive an asymptotic distribution for $\tilde{\eta}$, we shall establish that the following random matrices

$$\tilde{S}(n_T) = T^{1/2}\tilde{Q}(n_T)\tilde{\Omega}(n_T), \quad S_T = T^{1/2}Q\Omega_T, \tag{4.19}$$

are asymptotically equivalent.

Proposition 4.2 ASYMPTOTIC EQUIVALENCE. *Let* $\{Y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VARMA echelon representation given by* (2.11) - (2.15). *Then, under the assumptions* **2.1** *to* **2.4**, *the following equivalence holds*

$$\|\tilde{S}(n_T) - S_T\| = O_p\left(\frac{n_T^2}{T^{1/2}}\right).$$

Finally, we can give the asymptotic distribution of $\sqrt{T}(\tilde{\eta} - \eta)$.

Theorem 4.2 ASYMPTOTIC DISTRIBUTION OF TWO-STAGE ESTIMATOR. *Let* $\{Y_t : t \in \mathbb{Z}\}$ *be a k-dimensional stationary invertible stochastic process with the VARMA echelon representation given by* (2.11) - (2.15). *If the assumptions* **2.1** *to* **2.8** *are satisfied, then the asymptotic distribution of the estimator* $\tilde{\eta}$ *is the following :*

$$\sqrt{T}(\tilde{\eta} - \eta) \underset{T\to\infty}{\longrightarrow} \mathrm{N}\big[0,\, \Sigma_\eta\big]$$

where

$$\Sigma_\eta = Q\Sigma_{Xu}Q', \quad \Sigma_{Xu} = R'\big[\Sigma_u \otimes \Gamma\big]R, \tag{4.20}$$

$$Q = \big\{R'\Upsilon R\big\}^{-1}, \quad \Upsilon = I_k \otimes \Gamma, \quad \Gamma = \mathsf{E}\big[X_t X_t'\big], \tag{4.21}$$

$X_t = \big[V_t', Y_{t-1}', \ldots, Y_{t-\bar{p}}', u_{t-1}', \ldots, u_{t-\bar{p}}'\big]'$ *and* $V_t = Y_t - u_t$.

An important consequence of the above theorem is the fact that the asymptotic distribution of $\tilde{\eta}$ is the same as in the case where the innovations $u_{t-1}', \ldots, u_{t-\bar{p}}'$ are known rather than approximated by a long autoregression. Furthermore, the covariance matrix Σ_η can be consistently estimated by

$$\hat{\Sigma}_\eta = \tilde{Q}(n_T)\Big\{R'\big[\tilde{\Sigma}_u(n_T) \otimes \tilde{\Gamma}(n_T)\big]R\Big\}\tilde{Q}(n_T)', \tag{4.22}$$

where

$$\tilde{Q}(n_T) = \big\{R'\tilde{\Upsilon}(n_T)R\big\}^{-1}, \quad \tilde{\Upsilon}(n_T) = I_k \otimes \tilde{\Gamma}(n_T), \tag{4.23}$$

$$\tilde{\Gamma}(n_T) = \frac{1}{T}\sum_{t=1}^{T} \tilde{X}_t(n_T)\tilde{X}_t(n_T)', \quad \tilde{\Sigma}_u(n_T) = \frac{1}{T}\sum_{t=1}^{T} \tilde{u}_t(n_T)\tilde{u}_t(n_T)'. \tag{4.24}$$

Standard t and F-type tests may then be performed in the usual way.

5. Conclusion

In this paper, we have provided the asymptotic distribution of a simple two-stage estimator for VARMA models in echelon form. The estimator is consistent when the auxiliary long autoregression used to generate first step estimates of model innovations has an order n_T which increases to infinity at a rate inferior to T^{δ} with $0 < \delta_0 \leq \delta < 1/2$. Further, it has an asymptotic normal distribution provided n_T increases at a rate inferior to T^{δ} with $0 < \delta_0 \leq \delta < 1/4$. In the latter case, the asymptotic distribution is not affected by the fact that estimated lagged residuals are used.

The above results can be exploited in several ways. First, the two-stage estimates and the associated distributional theory can be directly used for inference on the VARMA model. In particular, they can be used for model selection purposes and to simplify the model (e.g., by eliminating insignificant coefficients). Second, two-stage estimates can be exploited to get more efficient estimators, such as ML estimators or estimators that are asymptotically to ML. This can be done, in particular, to achieve efficiency with Gaussian innovations. Note, however, that such gains of efficiency may not obtain if the innovations are not Gaussian. Thirdly, because of its simplicity, the two-stage linear estimator is especially well adapted for being used in the context of simulation-based inference procedures, such as bootstrap tests. Further, the asymptotic distribution provided above can be useful in order to improve the validity of the bootstrap. Several of these issues will be studied in a subsequent paper.

A. Appendix : Proofs

PROOF OF PROPOSITION **3.1** Let us write :

$$\left\|\tilde{\Sigma}_u(n_T) - \Sigma_u\right\| = \left\|\tilde{\Sigma}_u(n_T) - \hat{\Sigma}_T\right\| + \left\|\hat{\Sigma}_T - \Sigma_u\right\| \tag{A.1}$$

where

$$\hat{\Sigma}_T - \Sigma_u = \frac{1}{T}\sum_{t=1}^{T}\left[u_t u_t' - \Sigma_u\right], \tag{A.2}$$

$$\begin{aligned}
\tilde{\Sigma}_u(n_T) - \hat{\Sigma}_T &= \frac{1}{T}\sum_{t=1}^{T}\left\{\tilde{u}_t(n_T)\tilde{u}_t(n_T)' - u_t u_t'\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\{\left[\tilde{u}_t(n_T) - u_t\right]\tilde{u}_t(n_T)' + u_t\left[\tilde{u}_t(n_T) - u_t\right]'\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\{\left[\tilde{u}_t(n_T) - u_t\right]u_t' + u_t\left[\tilde{u}_t(n_T) - u_t\right]' + \left[\tilde{u}_t(n_T) - u_t\right]\left[\tilde{u}_t(n_T) - u_t\right]'\right\}.
\end{aligned} \tag{A.3}$$

By the assumptions **2.1** and **2.2**,

$$\hat{\Sigma}_T - \Sigma_u = \frac{1}{T}\sum_{t=1}^{T}\left[u_t u_t' - \Sigma_u\right] = O_p\left(\frac{1}{T}\right), \tag{A.4}$$

$$\frac{1}{T}\sum_{t=1}^{T}\|u_t\| = O_p(1), \quad \frac{1}{T}\sum_{t=1}^{T}\|u_t\|^2 = O_p(1). \tag{A.5}$$

Now

$$\tilde{u}_t(n_T) - u_t = \left[\Pi(n_T) - \tilde{\Pi}(n_T)\right]Y_t(n_T) + \sum_{\tau=n_T+1}^{\infty}\Pi_\tau Y_{t-\tau}, \tag{A.6}$$

hence

$$\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{u}_t(n_T) - u_t\right]u_t' = \left[\Pi(n_T) - \tilde{\Pi}(n_T)\right]C_{Yu}(n_T) + S_{Yu}(n_T) \tag{A.7}$$

where $Y_t(n_T) = \left[Y_{t-1}', \ldots, Y_{t-n_T}'\right]'$, and

$$C_{Yu}(n_T) = \frac{1}{T}\sum_{t=1}^{T}Y_t(n_T)u_t' = \left[C_{Yu}(1, T)', \ldots, C_{Yu}(n_T, T)'\right]', \tag{A.8}$$

$$C_{Yu}(\tau, T) = \frac{1}{T}\sum_{t=1}^{T}Y_{t-\tau}u_t', \tag{A.9}$$

$$S_{Yu}(n_T) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty} \Pi_\tau Y_{t-\tau}u_t'. \tag{A.10}$$

Using the fact that u_t is independent of X_t, $u_{t-1}, \ldots, u_1$, we see that

$$\begin{aligned}
\mathsf{E}\|C_{Yu}(\tau, T)\|^2 &= \mathsf{E}\big[C_{Yu}(\tau, T)'C_{Yu}(\tau, T)\big] = \frac{1}{T^2}\sum_{t=1}^{T}\mathsf{E}\big[\mathrm{tr}(u_tY_{t-\tau}'Y_{t-\tau}u_t')\big] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{tr}\big[\mathsf{E}(u_t'u_t)\mathsf{E}(Y_{t-\tau}'Y_{t-\tau})\big] = \frac{1}{T}\mathrm{tr}\big[\Sigma_u\big]\mathrm{tr}\big[\Gamma(0)\big], \qquad (A.11) \\
\mathsf{E}\big[S_{Yu}(n_T)\big] &= 0, \qquad (A.12)
\end{aligned}$$

where $\Gamma(0) = \mathsf{E}\big[Y_tY_t'\big]$, hence

$$\begin{aligned}
\mathsf{E}\|C_{Yu}(n_T)\|^2 &= \mathsf{E}\big[C_{Yu}(n_T)'C_{Yu}(n_T)\big] = \sum_{\tau=1}^{n_T}\mathsf{E}\|C_{Yu}(\tau, T)\|^2 \\
&= \frac{n_T}{T}\mathrm{tr}\big[\Sigma_u\big]\mathrm{tr}\big[\Gamma(0)\big], \qquad (A.13) \\
\sum_{\tau=1}^{n_T}\|C_{Yu}(\tau, T)\|^2 &= O_p\Big(\frac{n_T}{T}\Big), \qquad (A.14)
\end{aligned}$$

and

$$\Big\|\big[\tilde{\Pi}(n_T) - \Pi(n_T)\big]C_{Yu}(n_T)\Big\| \le \|\tilde{\Pi}(n_T) - \Pi(n_T)\|\|C_{Yu}(n_T)\| = O_p\Big(\frac{n_T}{T}\Big). \tag{A.15}$$

Using the stationarity of Y_t and (2.8), we have :

$$\begin{aligned}
\mathsf{E}\|S_{Yu}(n_T)\| &\le \mathsf{E}\Big\{\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\|Y_{t-\tau}\|\|u_t\|\Big)\Big\} \\
&\le \Big\{\mathsf{E}\|Y_t\|^2\Big\}^{1/2}\Big\{\mathsf{E}\|u_t\|^2\Big\}^{1/2}\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\Big) \\
&\le \Big\{\mathsf{E}\|Y_t\|^2\Big\}^{1/2}\Big\{\mathsf{E}\|u_t\|^2\Big\}^{1/2}\frac{C}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\rho^\tau\Big) \\
&\le \Big\{\mathsf{E}\|Y_t\|^2\Big\}^{1/2}\Big\{\mathsf{E}\|u_t\|^2\Big\}^{1/2}\frac{C}{T}\sum_{t=1}^{T}\Big(\frac{\rho^{n_T+1}}{1-\rho}\Big) \\
&= \Big\{\mathsf{E}\|Y_t\|^2\Big\}^{1/2}\Big\{\mathsf{E}\|u_t\|^2\Big\}^{1/2}\Big(\frac{C\rho}{1-\rho}\Big)\rho^{n_T} = O(\rho^{n_T}) \qquad (A.16)
\end{aligned}$$

hence

$$\|S_{Yu}(n_T)\| = O_p(\rho^{n_T}). \tag{A.17}$$

Consequently,

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T}u_t\left[\tilde{u}_t(n_T)-u_t\right]'\right\| &= \left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{u}_t(n_T)-u_t\right]u_t'\right\| \\ &\leq \left\|\left[\tilde{\Pi}(n_T)-\Pi(n_T)\right]C_{Yu}(n_T)\right\|+\left\|S_{Yu}(n_T)\right\| \\ &= O_p\left(\frac{n_T}{T}\right), \end{aligned} \tag{A.18}$$

and (3.14) is established. Finally,

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{u}_t(n_T)-u_t\right]\left[\tilde{u}_t(n_T)-u_t\right]'\right\| &\leq \frac{1}{T}\sum_{t=1}^{T}\left\|\left[\tilde{u}_t(n_T)-u_t\right]\left[\tilde{u}_t(n_T)-u_t\right]'\right\| \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{u}_t(n_T)-u_t\right\|^2 \end{aligned} \tag{A.19}$$

where

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{u}_t(n_T)-u_t\right\|^2 &\leq \frac{3}{T}\sum_{t=1}^{T}\left\{\left\|\tilde{\Pi}(n_T)-\Pi(n_T)\right\|^2\left\|Y_t(n_T)\right\|^2\right. \\ &\quad \left.+\left(\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\left\|Y_{t-\tau}\right\|\right)^2\right\} \\ &\leq 3\left\|\tilde{\Pi}(n_T)-\Pi(n_T)\right\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|Y_t(n_T)\right\|^2 \\ &\quad +\frac{3}{T}\sum_{t=1}^{T}\left(\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\left\|Y_{t-\tau}\right\|\right)^2. \end{aligned} \tag{A.20}$$

Since

$$\mathsf{E}\left[\frac{1}{T}\sum_{t=1}^{T}\left\|Y_t(n_T)\right\|^2\right]=\mathsf{E}\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{n_T}\left\|Y_{t-\tau}\right\|^2\right]=n_T\,\mathsf{E}\left\|Y_t\right\|^2, \tag{A.21}$$

we have

$$\frac{1}{T}\sum_{t=1}^{T}\left\|Y_t(n_T)\right\|^2=O_p(n_T). \tag{A.22}$$

Further,

$$\begin{aligned}\mathsf{E}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\left\|Y_{t-\tau}\right\|\right)\right] &= \mathsf{E}\left\|Y_t\right\|\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right) \\ &\leq \mathsf{E}\left\|Y_t\right\|\frac{C}{T}\sum_{t=1}^{T}\left(\frac{\rho^{n_T+1}}{1-\rho}\right)=\left(\frac{C\,\mathsf{E}\left\|Y_t\right\|\rho}{1-\rho}\right)\rho^{n_T} \end{aligned}$$

$$= O(\rho^{n_T}), \tag{A.23}$$

hence

$$\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\|Y_{t-\tau}\|\Big) = O_p(\rho^{n_T}), \tag{A.24}$$

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\|Y_{t-\tau}\|\Big)^2 &\le T\Big\{\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\|Y_{t-\tau}\|\Big)\Big\}^2 \\ &= O_p\Big(T\rho^{2n_T}\Big).\end{aligned} \tag{A.25}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\|\tilde{u}_t(n_T)-u_t\|^2 \le O_p\Big(\frac{n_T}{T}\Big)O_p(n_T)+O_p\Big(T\rho^{2n_T}\Big) = O_p\Big(\frac{n_T^2}{T}\Big), \tag{A.26}$$

$$\Big\|\frac{1}{T}\sum_{t=1}^{T}\big[\tilde{u}_t(n_T)-u_t\big]\big[\tilde{u}_t(n_T)-u_t\big]'\Big\| = O_p\Big(\frac{n_T^2}{T}\Big). \tag{A.27}$$

We can thus conclude that

$$\|\tilde{\Sigma}_u(n_T)-\hat{\Sigma}_T\| = O_p\Big(\frac{n_T}{T}\Big)+O_p\Big(\frac{n_T^2}{T}\Big) = O_p\Big(\frac{n_T^2}{T}\Big), \tag{A.28}$$

$$\|\tilde{\Sigma}_u(n_T)-\Sigma_u\| = O_p\Big(\frac{n_T^2}{T}\Big). \tag{A.29}$$

□

PROOF OF PROPOSITION **4.1** Using (4.2) and (4.8), we see that

$$\begin{aligned}\tilde{\Gamma}(n_T)-\Gamma_T &= \frac{1}{T}\sum_{t=1}^{T}\big[\tilde{X}_t(n_T)\tilde{X}_t(n_T)'-X_tX_t'\big] \\ &= \frac{1}{T}\sum_{t=1}^{T}\Big\{\big[\tilde{X}_t(n_T)-X_t\big]X_t'+X_t\big[\tilde{X}_t(n_T)-X_t\big]'\Big\} \\ &\quad+\frac{1}{T}\sum_{t=1}^{T}\Big\{\big[\tilde{X}_t(n_T)-X_t\big]\big[\tilde{X}_t(n_T)-X_t\big]'\Big\}\end{aligned} \tag{A.30}$$

hence, using the triangular and Cauchy-Schwarz inequalities,

$$\begin{aligned}\left\|\tilde{\Gamma}(n_T)-\Gamma_T\right\| &\leq 2\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|X_t\right\|^2\right\}^{1/2}\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{X}_t(n_T)-X_t\right\|^2\right\}^{1/2} \\ &\quad+\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{X}_t(n_T)-X_t\right\|^2 \\ &= 2\left\{\frac{1}{T}\left\|X(T)\right\|^2\right\}^{1/2}\left\{\frac{1}{T}\left\|\tilde{X}(n_T)-X(T)\right\|^2\right\}^{1/2} \\ &\quad+\frac{1}{T}\left\|\tilde{X}(n_T)-X(T)\right\|^2 \end{aligned} \tag{A.31}$$

where

$$\tilde{X}_t(n_T)-X_t=\begin{bmatrix} u_t-\tilde{u}_t(n_T) \\ 0 \\ \vdots \\ 0 \\ \tilde{u}_{t-1}(n_T)-u_{t-1} \\ \vdots \\ \tilde{u}_{t-\tilde{p}}(n_T)-u_{t-\tilde{p}} \end{bmatrix}, \tag{A.32}$$

$$\begin{aligned}\frac{1}{T}\left\|\tilde{X}(n_T)-X(T)\right\|^2 &= \frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{X}_t(n_T)-X_t\right\|^2 \\ &= \sum_{j=0}^{\tilde{p}}\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{u}_{t-j}(n_T)-u_{t-j}\right\|^2\right\}=O_p\left(\frac{n_T^2}{T}\right)\end{aligned} \tag{A.33}$$

and, by the stationarity assumption,

$$\frac{1}{T}\left\|X(T)\right\|^2=\frac{1}{T}\sum_{t=1}^{T}\left\|X_t\right\|^2=O_p(1). \tag{A.34}$$

It follows from the above orders that

$$\left\|\tilde{\Gamma}(n_T)-\Gamma_T\right\|=O_p\left(\frac{n_T}{T^{1/2}}\right). \tag{A.35}$$

Consequently, we have :

$$\begin{aligned}\|\tilde{\Upsilon}(n_T) - \Upsilon_T\| &= \left\|[I_k \otimes \tilde{\Gamma}(n_T)] - [I_k \otimes \Gamma_T]\right\| \\ &= \left\|I_k \otimes [\tilde{\Gamma}(n_T) - \Gamma_T]\right\| \\ &= k^{1/2}\|\tilde{\Gamma}(n_T) - \Gamma_T\| = O_p\left(\frac{n_T}{T^{1/2}}\right), \end{aligned} \tag{A.36}$$

$$\begin{aligned}\|\tilde{Q}(n_T)^{-1} - Q_T^{-1}\| &= \left\|R'[\tilde{\Upsilon}(n_T) - \Upsilon_T]R\right\| \\ &\leq \|R\|^2\|\tilde{\Upsilon}(n_T) - \Upsilon_T\| = O_p\left(\frac{n_T}{T^{1/2}}\right). \end{aligned} \tag{A.37}$$

Further, since

$$\|\tilde{Q}(n_T)^{-1} - Q^{-1}\| \leq \|\tilde{Q}(n_T)^{-1} - Q_T^{-1}\| + \|Q_T^{-1} - Q^{-1}\| \tag{A.38}$$

and

$$\begin{aligned}\|Q_T^{-1} - Q^{-1}\| &= \left\|R'[\Upsilon_T - \Upsilon]R\right\| \leq \|R\|^2\|\Upsilon_T - \Upsilon\| \\ &\leq \|R\|^2\left\|I_k \otimes [\Gamma_T - \Gamma]\right\| = k^{1/2}\|R\|^2\|\Gamma_T - \Gamma\| \\ &= k^{1/2}\|R\|^2\left\|\frac{1}{T}\sum_{t=1}^{T} X_t X_t' - \mathsf{E}[X_t X_t']\right\| = O_p\left(\frac{1}{T^{1/2}}\right), \end{aligned} \tag{A.39}$$

we have :

$$\|\tilde{Q}(n_T)^{-1} - Q^{-1}\| = O_p\left(\frac{n_T}{T^{1/2}}\right). \tag{A.40}$$

Finally, using the triangular inequality, we get :

$$\|\tilde{Q}(n_T)\| \leq \|\tilde{Q}(n_T) - Q\| + \|Q\|, \tag{A.41}$$

$$\begin{aligned}\|\tilde{Q}(n_T) - Q\| &= \left\|\tilde{Q}(n_T)[\tilde{Q}(n_T)^{-1} - Q^{-1}]Q\right\| \\ &\leq \|\tilde{Q}(n_T)\|\|\tilde{Q}(n_T)^{-1} - Q^{-1}\|\|Q\| \end{aligned}$$

$$\leq \left[\|\tilde{Q}(n_T) - Q\| + \|Q\|\right]\|\tilde{Q}(n_T)^{-1} - Q^{-1}\|\|Q\|, \tag{A.42}$$

hence, for $\|\tilde{Q}(n_T)^{-1} - Q^{-1}\|\|Q\| < 1$ (an event whose probability converges to 1 as $T \to \infty$)

$$\|\tilde{Q}(n_T) - Q\| \leq \frac{\|Q\|^2\|\tilde{Q}(n_T)^{-1} - Q^{-1}\|}{1 - \|\tilde{Q}(n_T)^{-1} - Q^{-1}\|\|Q\|} = O_p\left(\frac{n_T}{T^{1/2}}\right). \tag{A.43}$$

□

PROOF OF THEOREM **4.1** Recall that $\tilde{\eta} - \eta = \tilde{Q}(n_T)\tilde{\Omega}(n_T)$. Then, we have

$$\begin{aligned}\|\tilde{\eta} - \eta\| &\leq \|Q\|_1\|\Omega_T\| + \|\tilde{Q}(n_T) - Q\|_1\|\Omega_T\| + \|\tilde{Q}(n_T)\|_1\|\tilde{\Omega}(n_T) - \Omega_T\| \\ &\leq \|Q\|\|\Omega_T\| + \|\tilde{Q}(n_T) - Q\|\|\Omega_T\| + \|\tilde{Q}(n_T)\|\|\tilde{\Omega}(n_T) - \Omega_T\|. \end{aligned} \tag{A.44}$$

By Proposition **4.1**,

$$\|\tilde{Q}(n_T) - Q\| = O_p\left(\frac{n_T}{T^{1/2}}\right), \quad \|\tilde{Q}(n_T)\| = O_p(1). \tag{A.45}$$

Now

$$\Omega_T = \frac{1}{T}\sum_{t=1}^{T} R'[I_k \otimes X_t]\, u_t = R'\mathrm{vec}\left[\frac{1}{T}\sum_{t=1}^{T} X_t u_t'\right], \tag{A.46}$$

so that

$$\mathsf{E}\|\Omega_T\|^2 \leq \|R\|^2\mathsf{E}\|W_T\|^2 \tag{A.47}$$

where

$$W_T = \frac{1}{T}\sum_{t=1}^{T} X_t u_t'. \tag{A.48}$$

Then, using the fact that u_t is independent of X_t, $u_{t-1}, \ldots, u_1$,

$$\begin{aligned}\mathsf{E}\|W_T\|^2 &= \mathsf{E}\left[\mathrm{tr}\left(W_T W_T'\right)\right] \\ &= \frac{1}{T^2}\left\{\sum_{t=1}^{T}\mathsf{E}\left(\mathrm{tr}\left[X_t u_t' u_t X_t'\right]\right) + 2\sum_{t=1}^{T-1}\sum_{l=1}^{T-l}\mathsf{E}\left(\mathrm{tr}\left[X_t u_t' u_{t+l} X_{t+l}'\right]\right)\right\} \\ &= \frac{1}{T^2}\left\{\sum_{t=1}^{T}\mathsf{E}\left(\mathrm{tr}\left[u_t' u_t X_t' X_t\right]\right) + 2\sum_{t=1}^{T-1}\sum_{l=1}^{T-l}\mathsf{E}\left(\mathrm{tr}\left[u_{t+l} X_{t+l}' X_t u_t'\right]\right)\right\}\end{aligned}$$

$$= \frac{1}{T^2}\Big\{\sum_{t=1}^{T}\mathrm{tr}\Big[\mathsf{E}(u_t'u_t)\mathsf{E}(X_t'X_t)\Big] + 2\sum_{t=1}^{T-1}\sum_{l=1}^{T-l}\mathrm{tr}\Big[\mathsf{E}(u_{t+l})\mathsf{E}(X_{t+l}'X_tu_t')\Big]\Big\}$$

$$= \frac{1}{T^2}\Big\{\sum_{t=1}^{T}\mathrm{tr}\Big[\mathsf{E}(u_tu_t')\mathsf{E}(X_t'X_t)\Big]\Big\} = \frac{1}{T}\mathrm{tr}\big[\Sigma_u\big]\mathrm{tr}\big[\Gamma\big] \tag{A.49}$$

hence

$$\|W_T\| = O_p\Big(T^{-1/2}\Big), \quad \|\Omega_T\| = O_p\Big(T^{-1/2}\Big). \tag{A.50}$$

Now, consider the term $\|\tilde{\Omega}(n_T) - \Omega_T\|$. We have :

$$\begin{aligned}\tilde{\Omega}(n_T) - \Omega_T &= \frac{1}{T}\sum_{t=1}^{T} R'\Big\{\big[I_k \otimes \tilde{X}_t(n_T)\big]e_t(n_T) - \big[I_k \otimes X_t\big]u_t\Big\} \\ &= R'\mathrm{vec}\Big[\frac{1}{T}\sum_{t=1}^{T}\Big\{\tilde{X}_t(n_T)e_t(n_T)' - X_tu_t'\Big\}\Big] \\ &= R'\mathrm{vec}\big[\tilde{\Omega}_1(n_T) + \tilde{\Omega}_2(n_T)\big]\end{aligned} \tag{A.51}$$

where

$$\tilde{\Omega}_1(n_T) = \frac{1}{T}\sum_{t=1}^{T} X_t\big[e_t(n_T) - u_t\big]', \tag{A.52}$$

$$\tilde{\Omega}_2(n_T) = \frac{1}{T}\sum_{t=1}^{T}\big[\tilde{X}_t(n_T) - X_t\big]e_t(n_T)', \tag{A.53}$$

$$e_t(n_T) = \tilde{u}_t(n_T) + \sum_{j=0}^{\bar{p}}\Theta_j\big[u_{t-j} - \tilde{u}_{t-j}(n_T)\big]. \tag{A.54}$$

We can also write

$$e_t(n_T) - u_t = \sum_{j=0}^{\bar{p}}\bar{\Theta}_j\big[\tilde{u}_{t-j}(n_T) - u_{t-j}\big] \tag{A.55}$$

where $\bar{\Theta}_0 = I_k - \Theta_0$ and $\bar{\Theta}_j = -\Theta_j$, $j = 1, 2, \ldots, \bar{p}$, and

$$\begin{aligned}\tilde{u}_t(n_T) - u_t &= \big[\Pi(n_T) - \tilde{\Pi}(n_T)\big]Y_t(n_T) + \sum_{\tau=n_T+1}^{\infty}\Pi_\tau Y_{t-\tau} \\ &= \sum_{\tau=1}^{n_T}\big[\Pi_\tau - \tilde{\Pi}_\tau(n_T)\big]Y_{t-\tau} + \sum_{\tau=n_T+1}^{\infty}\Pi_\tau Y_{t-\tau},\end{aligned} \tag{A.56}$$

hence

$$\begin{aligned}
\tilde{\Omega}_1(n_T) &= \frac{1}{T}\sum_{t=1}^{T} X_t\big[e_t(n_T)-u_t\big]' \\
&= \sum_{j=0}^{\bar{p}}\Big(\frac{1}{T}\sum_{t=1}^{T}\Big\{\sum_{\tau=1}^{n_T} X_t Y'_{t-j-\tau}\big[\Pi_\tau-\tilde{\Pi}_\tau(n_T)\big]' + \sum_{\tau=n_T+1}^{\infty} X_t Y'_{t-j-\tau}\Pi'_\tau\Big\}\Big)\bar{\Theta}'_j \\
&= \sum_{j=0}^{\bar{p}}\Big\{\sum_{\tau=1}^{n_T}\Big(\frac{1}{T}\sum_{t=1}^{T} X_t Y'_{t-j-\tau}\Big)\big[\Pi_\tau-\tilde{\Pi}_\tau(n_T)\big]' + \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty} X_t Y'_{t-j-\tau}\Pi'_\tau\Big\}\bar{\Theta}'_j \\
&= \tilde{\Omega}_{11}(n_T)+\tilde{\Omega}_{12}(n_T) \qquad \text{(A.57)}
\end{aligned}$$

where

$$\tilde{\Omega}_{11}(n_T) = \sum_{j=0}^{\bar{p}}\Big\{\sum_{\tau=1}^{n_T}\tilde{\Gamma}_{j+\tau}(n_T)\big[\Pi_\tau-\tilde{\Pi}_\tau(n_T)\big]'\Big\}\bar{\Theta}'_j, \qquad \text{(A.58)}$$

$$\tilde{\Gamma}_{j+\tau}(n_T) = \frac{1}{T}\sum_{t=1}^{T} X_t Y'_{t-j-\tau}, \qquad \text{(A.59)}$$

$$\tilde{\Omega}_{12}(n_T) = \sum_{j=0}^{\bar{p}}\Big\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty} X_t Y'_{t-j-\tau}\Pi'_\tau\Big\}\bar{\Theta}'_j. \qquad \text{(A.60)}$$

Now, using the linearity and the VARMA structure of Y_t, it is easy to see that

$$\mathsf{E}\|\tilde{\Gamma}_{j+\tau}(n_T)\|^2 \le \frac{1}{T}C_1\rho_1^{j+\tau} \qquad \text{(A.61)}$$

for some constants $C_1 > 0$ and $0 < \rho_1 < 1$, hence

$$\mathsf{E}\Big[\sum_{\tau=1}^{n_T}\|\tilde{\Gamma}_{j+\tau}(n_T)\|^2\Big] \le \frac{C_1}{T}\Big(\sum_{\tau=1}^{n_T}\rho_1^{j+\tau}\Big) \le \frac{1}{T}\Big(\frac{C_1}{1-\rho_1}\Big) = O_p\Big(\frac{1}{T}\Big). \qquad \text{(A.62)}$$

Thus

$$\begin{aligned}
\|\tilde{\Omega}_{11}(n_T)\| &\le \sum_{j=0}^{\bar{p}}\Big\{\sum_{\tau=1}^{n_T}\|\tilde{\Gamma}_{j+\tau}(n_T)\|\|\Pi_\tau-\tilde{\Pi}_\tau(n_T)\|\Big\}\|\bar{\Theta}_j\| \\
&\le \sum_{j=0}^{\bar{p}}\Big\{\sum_{\tau=1}^{n_T}\|\tilde{\Gamma}_{j+\tau}(n_T)\|^2\Big\}^{1/2}\Big\{\sum_{\tau=1}^{n_T}\|\Pi_\tau-\tilde{\Pi}_\tau(n_T)\|^2\Big\}^{1/2}\|\bar{\Theta}_j\| \\
&\le \sum_{j=0}^{\bar{p}}\Big\{\sum_{\tau=1}^{n_T}\|\tilde{\Gamma}_{j+\tau}(n_T)\|^2\Big\}^{1/2}\|\tilde{\Pi}(n_T)-\Pi(n_T)\|\|\bar{\Theta}_j\|
\end{aligned}$$

$$= O_p\Big(\frac{n_T^{1/2}}{T}\Big), \tag{A.63}$$

while

$$\begin{aligned}
\mathsf{E}\|\tilde{\Omega}_{12}(n_T)\| &\leq \sum_{j=0}^{\bar{p}}\Big\{\mathsf{E}\Big[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\|X_t\|\|Y_{t-j-\tau}\|\|\Pi_\tau\|\Big]\Big\}\|\bar{\Theta}_j\| \\
&\leq \sum_{j=0}^{\bar{p}}\Big\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\mathsf{E}\Big[\|X_t\|\|Y_{t-j-\tau}\|\Big]\Big\}\|\bar{\Theta}_j\| \\
&\leq \sum_{j=0}^{\bar{p}}\Big\{\Big(\mathsf{E}\|X_t\|^2\Big)\Big(\mathsf{E}\|Y_t\|^2\Big)\Big\}^{1/2}\Big\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\Big\}\|\bar{\Theta}_j\| \\
&= O_p(\rho^{n_T}),
\end{aligned} \tag{A.64}$$

hence $\|\tilde{\Omega}_{12}(n_T)\| = O_p(\rho^{n_T})$ and

$$\|\tilde{\Omega}_1(n_T)\| \leq \|\tilde{\Omega}_{11}(n_T)\| + \|\tilde{\Omega}_{12}(n_T)\| = O_p\Big(\frac{n_T^{1/2}}{T}\Big). \tag{A.65}$$

Now, using (A.55), $\tilde{\Omega}_2(n_T)$ can be decomposed as :

$$\tilde{\Omega}_2(n_T) = \tilde{\Omega}_{21}(n_T) + \tilde{\Omega}_{22}(n_T) \tag{A.66}$$

where

$$\tilde{\Omega}_{21}(n_T) = \frac{1}{T}\sum_{t=1}^{T}\big[\tilde{X}_t(n_T) - X_t\big]u_t', \tag{A.67}$$

$$\tilde{\Omega}_{22}(n_T) = \sum_{j=0}^{\bar{p}}\Big\{\frac{1}{T}\sum_{t=1}^{T}\big[\tilde{X}_t(n_T) - X_t\big]\big[\tilde{u}_{t-j}(n_T) - u_{t-j}\big]'\Big\}\bar{\Theta}_j'. \tag{A.68}$$

Now, in view of (A.32), consider the variables :

$$\begin{aligned}
C_i(n_T) &= \frac{1}{T}\sum_{t=1}^{T}\big[\tilde{u}_{t-i}(n_T) - u_{t-i}\big]u_t' \\
&= \sum_{\tau=1}^{n_T}\big[\Pi_\tau - \tilde{\Pi}_\tau(n_T)\big]\Big(\frac{1}{T}\sum_{t=1}^{T}Y_{t-i-\tau}u_t'\Big) + \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\Pi_\tau Y_{t-i-\tau}u_t',
\end{aligned} \tag{A.69}$$

$$C_{ij}(n_T) = \frac{1}{T}\sum_{t=1}^{T}\big[\tilde{u}_{t-i}(n_T) - u_{t-i}\big]\big[\tilde{u}_{t-j}(n_T) - u_{t-j}\big]', \tag{A.70}$$

for $i = 0, 1, \ldots, \bar{p}$. We have :

$$\mathsf{E}\left\|\frac{1}{T}\sum_{t=1}^{T} Y_{t-i-\tau}u_t'\right\|^2 = \frac{1}{T^2}\sum_{t=1}^{T}\mathsf{E}\Big(\mathrm{tr}\big[Y_{t-i-\tau}u_t'u_tY_{t-i-\tau}'\big]\Big) = \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{tr}\Big[\mathsf{E}\big(u_t'u_t\big)\mathsf{E}\big(Y_{t-i-\tau}'Y_{t-i-\tau}\big)\Big]$$

$$= \frac{1}{T}\mathrm{tr}\big[\Sigma_u\big]\mathrm{tr}\big[\Gamma(0)\big] \tag{A.71}$$

where $\Gamma(0) = \mathsf{E}\big[Y_tY_t'\big]$, hence

$$\sum_{\tau=1}^{n_T}\mathsf{E}\left\|\frac{1}{T}\sum_{t=1}^{T} Y_{t-i-\tau}u_t'\right\|^2 = \frac{n_T}{T}\mathrm{tr}\big[\Sigma_u\big]\mathrm{tr}\big[\Gamma(0)\big], \tag{A.72}$$

$$\sum_{\tau=1}^{n_T}\left\|\frac{1}{T}\sum_{t=1}^{T} Y_{t-i-\tau}u_t'\right\|^2 = O_p\Big(\frac{n_T}{T}\Big), \tag{A.73}$$

and

$$\begin{aligned}
\big\|C_i(n_T)\big\| &\leq \sum_{\tau=1}^{n_T}\big\|\Pi_\tau - \tilde{\Pi}_\tau(n_T)\big\|\left\|\frac{1}{T}\sum_{t=1}^{T} Y_{t-i-\tau}u_t'\right\| \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\big\|\Pi_\tau\big\|\big\|Y_{t-i-\tau}\big\|\big\|u_t\big\| \\
&\leq \left\{\sum_{\tau=1}^{n_T}\big\|\Pi_\tau - \tilde{\Pi}_\tau(n_T)\big\|^2\right\}^{1/2}\left\{\sum_{\tau=1}^{n_T}\left\|\frac{1}{T}\sum_{t=1}^{T} Y_{t-i-\tau}u_t'\right\|^2\right\}^{1/2} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\big\|\Pi_\tau\big\|\big\|Y_{t-i-\tau}\big\|\big\|u_t\big\| \\
&= \big\|\tilde{\Pi}(n_T) - \Pi(n_T)\big\|\left\{\sum_{\tau=1}^{n_T}\left\|\frac{1}{T}\sum_{t=1}^{T} Y_{t-i-\tau}u_t'\right\|^2\right\}^{1/2} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\big\|\Pi_\tau\big\|\big\|Y_{t-i-\tau}\big\|\big\|u_t\big\| \\
&= O_p\Big(\frac{n_T}{T}\Big).
\end{aligned} \tag{A.74}$$

Further,

$$\begin{aligned}
\big\|C_{ij}(n_T)\big\| &\leq \frac{1}{T}\sum_{t=1}^{T}\big\|\tilde{u}_{t-i}(n_T) - u_{t-i}\big\|\big\|\tilde{u}_{t-j}(n_T) - u_{t-j}\big\| \\
&\leq \left\{\frac{1}{T}\sum_{t=1}^{T}\big\|\tilde{u}_{t-i}(n_T) - u_{t-i}\big\|^2\right\}^{1/2}\left\{\frac{1}{T}\sum_{t=1}^{T}\big\|\tilde{u}_{t-j}(n_T) - u_{t-j}\big\|^2\right\}^{1/2} \\
&= O_p\Big(\frac{n_T^2}{T}\Big).
\end{aligned} \tag{A.75}$$

Thus

$$\|\tilde{\Omega}_{21}(n_T)\| = O_p\Big(\frac{n_T}{T}\Big), \quad \|\tilde{\Omega}_{22}(n_T)\| = O_p\Big(\frac{n_T^2}{T}\Big), \tag{A.76}$$

hence

$$\|\tilde{\Omega}_2(n_T)\| \leq \|\tilde{\Omega}_{21}(n_T)\| + \|\tilde{\Omega}_{22}(n_T)\| = O_p\Big(\frac{n_T^2}{T}\Big), \tag{A.77}$$

$$\begin{aligned} \|\tilde{\Omega}(n_T) - \Omega_T\| &\leq \|R\|\Big\{\|\tilde{\Omega}_1(n_T)\| + \|\tilde{\Omega}_2(n_T)\|\Big\} \\ &= O_p\Big(\frac{n_T^{1/2}}{T}\Big) + O_p\Big(\frac{n_T^2}{T}\Big) = O_p\Big(\frac{n_T^2}{T}\Big). \end{aligned} \tag{A.78}$$

Consequently,

$$\begin{aligned} \|\tilde{\eta} - \eta\| &\leq O_p\Big(\frac{1}{T^{1/2}}\Big) + O_p\Big(\frac{n_T}{T}\Big) + O_p\Big(\frac{n_T^2}{T}\Big) \\ &= O_p\Big(\frac{1}{T^{1/2}}\Big) + O_p\Big(\frac{n_T^2}{T}\Big) = o_p(1). \end{aligned} \tag{A.79}$$

If furthermore $n_T^4/T \longrightarrow 0$ as $T \to \infty$, the latter reduces to

$$\|\tilde{\eta} - \eta\| = O_p\Big(\frac{1}{T^{1/2}}\Big). \tag{A.80}$$

□

PROOF OF PROPOSITION **4.2** We have :

$$\begin{aligned} \|\tilde{S}(n_T) - S_T\| &= T^{1/2}\|\bar{Q}(n_T)\tilde{\Omega}(n_T) - Q\Omega_T\| \\ &\leq T^{1/2}\|\bar{Q}(n_T)\|\|\tilde{\Omega}(n_T) - \Omega_T\| + T^{1/2}\|\bar{Q}(n_T) - Q\|\|\Omega_T\|. \end{aligned} \tag{A.81}$$

By Proposition **4.1** and Theorem **4.1**, the following orders hold :

$$\|\bar{Q}(n_T) - Q\| = O_p\Big(\frac{n_T}{T^{1/2}}\Big), \quad \|\bar{Q}(n_T)\| = O_p(1), \tag{A.82}$$

$$\|\tilde{\Omega}(n_T) - \Omega_T\| \;=\; O_p\Big(\frac{n_T^2}{T}\Big), \quad \|\Omega_T\| = O_p\Big(\frac{1}{T^{1/2}}\Big). \tag{A.83}$$

Therefore,

$$\|\tilde{S}(n_T) - S_T\| = O_p\Big(\frac{n_T^2}{T^{1/2}}\Big). \tag{A.84}$$

□

PROOF OF THEOREM **4.2** By the standard central limit theorem for stationary processes [see Anderson (1971, section 7.7), Lewis and Reinsel (1985, section 2)] and under the assumption of independence between u_t and X_t, we have :

$$T^{1/2}\Omega_T = \frac{1}{T^{1/2}}\sum_{t=1}^{T} R'[I_k \otimes X_t]u_t = \frac{1}{T^{1/2}}\sum_{t=1}^{T} R'[u_t \otimes X_t] \underset{T\to\infty}{\longrightarrow} \mathrm{N}[0,\, \Sigma_{Xu}] \tag{A.85}$$

where

$$\begin{aligned} \Sigma_{Xu} &= \mathsf{E}\Big\{R'[u_t \otimes X_t][u_t \otimes X_t]'R\Big\} = \mathsf{E}\Big\{R'[u_t u_t' \otimes X_t X_t']R\Big\} \\ &= R'\Big(\mathsf{E}[u_t u_t'] \otimes \mathsf{E}[X_t X_t']\Big)R = R'[\Sigma_u \otimes \Gamma]R. \end{aligned} \tag{A.86}$$

Then

$$S_T = T^{1/2}Q\Omega_T \underset{T\to\infty}{\longrightarrow} \mathrm{N}[0,\, \Sigma_\eta] \tag{A.87}$$

where

$$\Sigma_\eta = Q\Sigma_{Xu}Q'. \tag{A.88}$$

Finally, by Proposition **4.2**, we can conclude that

$$\sqrt{T}(\tilde{\eta} - \eta) = \tilde{S}(n_T) \underset{T\to\infty}{\longrightarrow} \mathrm{N}[0,\, \Sigma_\eta]. \tag{A.89}$$

□

Chapter 3

Simplified order selection and efficient linear estimation for VARMA models with a macroeconomic application

1. Introduction

Modelling multivariate time series using vector autoregressive (VAR) models has received considerable attention, especially in econometrics ; see Lütkepohl (1991, 2001, 2005), Hamilton (1994, Chapter 11) and Dhrymes (1998). This popularity is due to the fact that such models are easy to estimate and can account for relatively complex dynamic phenomena. However, VAR models often require very large numbers of parameters in order to obtain good fits. Further, the VAR specification is not invariant to many basic linear transformations. For example, instead of satisfying a VAR scheme, subvectors follow vector autoregressive moving average (VARMA) processes. Temporal and contemporaneous aggregation lead to mixed VARMA models [see Lütkepohl (1987)]. Similarly, trend and seasonal adjustment also lead to models outside the VAR class [Maravall (1993)].

The VARMA structure includes VAR models as a special case, and can reproduce in a parsimonious way a much wider class of autocovariances and data generating processes (DGP). So they can lead to improvements in estimation and forecasting. VARMA modelling has been proposed a long time ago [see Hillmer and Tiao (1979), Tiao and Box (1981), Lütkepohl (1991), Boudjellaba, Dufour, and Roy (1992, 1994), Reinsel (1993, 1997)], but it has received little attention in practice. Although building VARMA models remains similar to the procedure associated with the univariate case, the task is compounded by the multivariate nature of the data.

At the specification level, several procedures which ensure a unique parameterization have been proposed ; see Hannan (1969b, 1970, 1971, 1976b, 1979, 1980a, 1981), Deistler and Hannan (1981), Deistler (1983), Hannan and Deistler (1988, Chapter 2), Lütkepohl (1991, Chapter 7) and Reinsel (1997, Chapter 3). In view of achieving both a parsimonious parameterization and efficiency, several methods have been considered. The main ones include : (1) techniques based on canonical analysis [Akaike (1974a, 1975, 1976), Cooper and Wood (1982), Tiao and Tsay (1985, 1989), Tsay and Tiao (1985), Tsay (1989a) and Paparoditis and Streitberg (1991)] ; (2) the Kronecker index approach which specifies an echelon-form VARMA representation [Deistler and

Hannan (1981), Hannan and Kavalieris (1984b), Solo (1986), Tsay (1989b), Nsiri and Roy (1992, 1996), Poskitt (1987, 1992), Lütkepohl and Poskitt (1996) and Bartel and Lütkepohl (1998)]; (3) the scalar-component model (SCM) approach [Tiao and Tsay (1989) and Tsay (1989b, 1991)]. In practice, however, the SCM and the canonical correlation techniques are complex and may lead to computational difficulties as they often involve the evaluation of a large number of eigenvalues. Furthermore, as pointed out by Lütkepohl and Poskitt (1996), consistency results for the estimation of the Kronecker indices by the canonical correlation approach do not appear to be available. That is why we shall use below model information criteria for specification purposes.

Once an identifiable specification has been formulated, different estimation methods have been considered. Methods based on Fourier transformations, recursive maximum likelihood (ML) and M-estimates have been proposed by Hannan (1969a, 1980b) and Kreiss (1985, 1987), respectively. But the most widely studied estimation method is ML for independent and identically distributed (i.i.d.) Gaussian innovations; see Newbold (1974), Box and Jenkins (1976), Hillmer and Tiao (1979), Nicholls and Hall (1979, 1980), Hannan, Kavalieris, and Mackisack (1986), Kohn (1981), Tiao and Box (1981), Solo (1984), Shea (1989), Mauricio (2002), and the review of Mélard, Roy, and Saidi (2002). However, maximizing the exact likelihood in stationary invertible VARMA models is computationally burdensome since for each autoregressive and moving average orders (say p and q) a non-quadratic optimization with respect to inequality constraints must be performed using iterative algorithms. As noted by Tiao and Box (1981), it is much easier to maximize a conditional likelihood, although in higher dimensional systems numerical problems still occur due to the lack of suitable initial values, even with known (p, q). Further, with weak white noise innovations, quasi-maximum likelihood estimates may not be consistent.

From the viewpoint of making VARMA modelling practical, one should have estimation methods that are both quick and simple to implement with standard software. Another reason for putting a premium on such estimation methods is that large-sample distributional theory tends to be quite unreliable in high-dimensional dynamic

models, so that tests and confidence sets based on asymptotic approximations are also unreliable. This suggests that simulation-based procedures - for example, bootstrap techniques - should be used. However, simulation may be impractical if calculating the estimator is difficult or time consuming.

In the univariate case, Hannan and Rissanen (1982) have proposed a recursive method which requires only linear regressions; see also Durbin (1960), Hannan and Kavalieris (1984a), Zhao-Guo (1985), Hannan, Kavalieris, and Mackisack (1986), Poskitt (1987), Koreisha and Pukkila (1990a, 1990b, 1995), Pukkila, Koreisha, and Kallinen (1990), Allende and Heiler (1992), Galbraith and Zinde-Walsh (1994, 1997) and Kavalieris, Hannan, and Salau (2003). This approach is based on estimating (by least squares) the innovations of the process through a long autoregression; then the VARMA parameters are estimated using the residuals from the long autoregression as regressors. Thereafter, new residuals are filtered and a linear regression using transformed variables is performed in order to achieve efficiency in the case of Gaussian innovations.

These methods have been extended to VARMA models; see Hannan and Kavalieris (1984b, 1986), Hannan and Deistler (1988), Koreisha and Pukkila (1989), Huang and Guo (1990), Reinsel, Basu, and Yap (1992), Poskitt (1992), Poskitt and Lütkepohl (1995), Lütkepohl and Poskitt (1996), Lütkepohl and Claessen (1997), De Frutos and Serrano (2002) and Dufour and Jouini (2005). It is worth noting that this linear estimation method (in its first two steps) has been introduced for model selection and getting initial values. After that, using other estimation procedures, such as ML, is typically suggested.

In this paper, we propose a general three-step linear estimation procedure for estimating stationary invertible VARMA models in echelon form. This approach can be easily adapted to VARMAX models and extended to integrated and cointegrated VARMA models as well. The estimation method focuses on the echelon form, since the latter tends to deliver relatively parsimonious parameterizations. But our procedure remains applicable to other identifying procedures such as final equations or any other restricted models for inference purposes.

In contrast with the above mentioned work on linear estimation methods for ARMA and VARMA models [with the exception of Bartel and Lütkepohl (1998)], we allow for the presence of intercepts. Further, we provide a general standard form for the parameter estimates, which is much easier to apply than the one considered by Hannan and Kavalieris (1984b). To do this, we extend the results developed in Dufour and Jouini (2005) to a generalized two-step estimation method and we derive the asymptotic distribution of the corresponding GLS estimators. We also provide a simple theoretical justification for the third-stage estimates proposed by Hannan and Kavalieris (1984b) - which they do not supply - and we show that such estimates correspond to a one-step iteration from the scoring algorithm, starting with the second-stage estimates as initial values. In addition, we prove under general conditions that these estimates have the same asymptotic distribution as the ML estimator in the case of Gaussian innovations. Unlike Reinsel, Basu, and Yap (1992), where Gaussian innovations are assumed and identification issues are not considered, we derive the asymptotic distribution of our third-stage linear estimators with strong white noise. Although our third-stage estimation procedure is equivalent to that of Hannan and Kavalieris (1984b), the estimates of the asymptotic covariances that we give for the third as well the second-stage estimators are relatively simple and easy to use.

Besides results on estimation, we propose a simplified order selection procedure to identify the Kronecker indices associated with the echelon-form VARMA representation. The procedure rests on the ability to determine the implied restriction matrix for all possible sets of Kronecker indices for any given dimension of the VARMA system. Indeed, this task is quite hard and could be an important reason for the lack of attraction of the echelon form in practice. In this paper, we develop an algorithm which solves this problem. We also supply an algorithm to formulate the implied restriction matrices corresponding to each equation [see Dufour and Jouini (2005)]. Hence, the identification of dynamic indices is much easier, and our procedure tends to reduce overparameterization and to improve accuracy - over the methods proposed by Hannan and Kavalieris (1984b) and Poskitt (1992). We propose information criteria (for the first two steps) and shortcuts that yield strongly consistent estimates of

the Kronecker indices. Unlike the Hannan and Kavalieris (1984b) and Poskitt (1992) approaches, which seem to work poorly in estimating the Kronecker indices, we show by simulation that our procedure behaves well and yields more accurate estimates than those proposed by Hannan and Kavalieris (1984b).

The paper proceeds as follows. In section **2,** we give the background model where an echelon VARMA representation is used to ensure a unique parameterization. Section **3** describes the generalized two-step linear estimation procedure (which allows for intercepts) and discusses the estimator properties such as convergence and asymptotic distribution. In section **4,** we provide a heuristic derivation of the third-stage estimators, and we demonstrate the asymptotic efficiency of these estimates when innovations are i.i.d. Gaussian. Section **5** discusses model selection and shortcuts to obtain the Kronecker indices. Further, as the estimated models may be nonstationary or noninvertible, we also provide an algorithm that ensures such constraints. Section **6** evaluates the proposed technique through a simulation study. Section **7** illustrates the method by testing for the long-run neutrality of monetary policy, using U.S. data. We conclude in section **8**. The proofs of the lemmas, propositions and theorems are supplied in Appendix **A**.

2. Framework

We consider a k-dimensional stochastic process of the autoregressive moving-average (VARMA) type with order (p, q). We first define the standard VARMA representation, which may involve identification problems. Then, among the representations which ensure parameter uniqueness in VARMA models, we proceed with the echelon form. Finally, we formulate the basic regularity assumptions we shall consider in the sequel.

2.1. Standard form

Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional random vector process with the VARMA representation

$$y_t = \mu_A + \sum_{i=1}^{p} A_i y_{t-i} + u_t + \sum_{j=1}^{q} B_j u_{t-j}, \tag{2.1}$$

where $y_t = (y_{1,t}, \ldots, y_{k,t})'$, $\mu_A = A(1)\,\mu_y$, $A(1) = I_k - \sum_{i=1}^{p} A_i$, $\mu_y = \mathsf{E}(y_t)$, p and q are non-negative integers (respectively, the autoregressive and moving average orders), A_i and B_j are $k \times k$ fixed coefficient matrices, $\{u_t : t \in \mathbb{Z}\} \sim \mathrm{WN}(0, \Sigma_u)$, i.e. u_t is a (second order) white noise process, such that $\Sigma_u = \mathsf{E}(u_t u_t')$, where Σ_u is a $k \times k$ positive definite symmetric matrix. Under stationarity and invertibility conditions the coefficients A_i and B_j satisfy the constraints $\det\{A(z)\} \neq 0$ and $\det\{B(z)\} \neq 0$ for all $|z| \leq 1$, where z is a complex number, $A(z) = I_k - \sum_{i=1}^{p} A_i z^i$ and $B(z) = I_k + \sum_{j=1}^{q} B_j z^j$. Then y_t has the following infinite-order autoregressive and moving average representations :

$$y_t = \mu_\Pi + \sum_{\tau=1}^{\infty} \Pi_\tau y_{t-\tau} + u_t, \tag{2.2}$$

$$y_t = \mu_y + u_t + \sum_{v=1}^{\infty} \Psi_v u_{t-v} \tag{2.3}$$

where

$$\Pi(z) = B(z)^{-1} A(z) = I_k - \sum_{\tau=1}^{\infty} \Pi_\tau z^\tau, \tag{2.4}$$

$$\Psi(z) = A(z)^{-1} B(z) = I_k + \sum_{v=1}^{\infty} \Psi_v z^v, \tag{2.5}$$

$\det\{\Pi(z)\} \neq 0$ and $\det\{\Psi(z)\} \neq 0$ for all $|z| \leq 1$, and $\mu_\Pi = \Pi(1)\,\mu_y$ with $\Pi(1) = I_k - \sum_{\tau=1}^{\infty} \Pi_\tau$. Moreover, we can find real constants $C > 0$ and $\rho \in (0,1)$ such that

$$\|\Pi_\tau\| \leq C\rho^\tau, \quad \|\Psi_v\| \leq C\rho^v, \tag{2.6}$$

hence $\sum_{\tau=1}^{\infty} \|\Pi_\tau\| < \infty$ and $\sum_{v=1}^{\infty} \|\Psi_v\| < \infty$, where $\|.\|$ stands for Schur norm [see Horn and Johnson (1985, Section 5.6)], *i.e.* $\|M\|^2 = \mathrm{tr}\,[M'M]$ for any matrix M.

2.2. Echelon form

The standard VARMA(p,q) representation (2.1) is not unique. The coefficient matrices A_i and B_j are not uniquely determined by the covariance structure (although the coefficients Π_τ and Ψ_v typically are). To ensure a unique parameterization of (2.1) we consider the stationary invertible VARMA(p,q) process in echelon form

$$\Phi(L)\left(y_t - \mu_y\right) = \Theta(L)\, u_t \tag{2.7}$$

or, equivalently,

$$\Phi(L)\, y_t = \mu_\Phi + \Theta(L)\, u_t \tag{2.8}$$

where $\Phi(L) = \Phi_0 - \sum_{i=1}^{\bar{p}} \Phi_i L^i$, $\Theta(L) = \Theta_0 + \sum_{j=1}^{\bar{p}} \Theta_j L^j$, L denotes the lag operator, $\mu_\Phi = \Phi(1)\ \mu_y$, $\bar{p} = \max(p,q)$, $\Theta_0 = \Phi_0$, and Φ_0 is a lower-triangular matrix whose diagonal elements are all equal to one. The VARMA representation (2.8) is in echelon form if $\Phi(L) = [\phi_{lm}(L)]_{l,m=1,\ldots,k}$ and $\Theta(L) = [\theta_{lm}(L)]_{l,m=1,\ldots,k}$ satisfy the following conditions : given a vector of orders $P = (p_1, \ldots, p_k)'$ called the Kronecker indices, the operators $\phi_{lm}(L)$ and $\theta_{lm}(L)$ on any given row l of $\Phi(L)$ and $\Theta(L)$ have the same degree p_l and

$$\begin{aligned} \phi_{lm}(L) &= 1 - \sum_{i=1}^{p_l} \phi_{ll,i} L^i \qquad \text{if } l = m, \\ &= -\sum_{i=p_l - p_{lm}+1}^{p_l} \phi_{lm,i} L^i \quad \text{if } l \neq m, \end{aligned} \tag{2.9}$$

$$\theta_{lm}(L) = \sum_{j=0}^{p_l} \theta_{lm,j} L^j, \text{ with } \Theta_0 = \Phi_0, \tag{2.10}$$

for $l, m = 1, \ldots, k$, where

$$\begin{aligned} p_{lm} &= \min(p_l + 1, p_m) \quad \text{for } l \geq m, \\ &= \min(p_l, p_m) \qquad\quad \text{for } l < m. \end{aligned} \tag{2.11}$$

Note that $p_{ll} = p_l$ is the number of free coefficients on the l-th diagonal element of $\Phi(L)$ as well the order of the polynomials on the corresponding row of $\Theta(L)$, while p_{lm} specifies the number of free coefficients in the operator $\phi_{lm}(L)$ for $l \neq m$. The sum of the Kronecker indices $\sum_{l=1}^{k} p_l$ is called the McMillan degree and the matrix formed by the Kronecker indices associated with the model is $\bar{P} = [p_{lm}]_{l,m=1,\ldots,k}$. This leads to $\sum_{l=1}^{k}\sum_{m=1}^{k} p_{lm}$ autoregressive and $k\sum_{l=1}^{k} p_l$ moving average free coefficients, respectively. Clearly, for the echelon form VARMA orders, we have $\bar{p} = \max(p_1, \ldots, p_k)$. Moreover, this identified parameterization of VARMA(p, q) models, that we refer, henceforward, as VARMA$(p_1, \ldots, p_k)$, ensures the uniqueness of left-coprime operators $\Phi(L)$ and $\Theta(L)$. Among other identifiable parameterizations, such as the final equations form, the echelon form has been preferred for parsimony and gain efficiency criteria. For proofs of the uniqueness of the echelon form and other identification conditions, the reader should consult Hannan (1969b, 1970, 1976b, 1979), Deistler and Hannan (1981), Hannan and Deistler (1988), and Lütkepohl (1991, Chapter 7).

The stationarity and invertibility conditions in (2.8) are : $\det\{\Phi(z)\} \neq 0$ and $\det\{\Theta(z)\} \neq 0$ for all $|z| \leq 1$, where $\Phi(z) = \Phi_0 - \sum_{i=1}^{\bar{p}} \Phi_i z^i$, $\Theta(z) = \Theta_0 + \sum_{j=1}^{\bar{p}} \Theta_j z^j$, with $\Pi(z) = \Theta(z)^{-1}\Phi(z)$ and $\Psi(z) = \Phi(z)^{-1}\Theta(z)$. Let also

$$\Theta(z)^{-1} = \sum_{\tau=0}^{\infty} \Lambda_\tau(\eta) z^\tau \tag{2.12}$$

where by invertibility $\|\Lambda_\tau(\eta)\| \leq C\rho^\tau$, $\sum_{\tau=0}^{\infty} \|\Lambda_\tau(\eta)\| < \infty$, and η is the vector of all free varying parameters implied by the echelon form, as it will be specified further. Now, set $v_t = y_t - u_t$. We see easily that

$$v_t = \Phi_0^{-1}\Big[\mu_\Phi + \sum_{i=1}^{\bar{p}} \Phi_i y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j u_{t-j}\Big]. \tag{2.13}$$

Obviously, v_t is uncorrelated with the error term u_t and (2.8) takes the form

$$y_t = \mu_\Phi + \big(I_k - \Phi_0\big)v_t + \sum_{i=1}^{\bar{p}} \Phi_i y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j u_{t-j} + u_t. \tag{2.14}$$

Set

$$\beta = \operatorname{vec}\big[\mu_\Phi, I_k - \Phi_0, \Phi_1, \ldots, \Phi_{\bar{p}}, \Theta_1, \ldots, \Theta_{\bar{p}}\big], \tag{2.15}$$

$$X_t = \big[1, v_t', y_{t-1}', \ldots, y_{t-\bar{p}}', u_{t-1}', \ldots, u_{t-\bar{p}}'\big]' \tag{2.16}$$

where β and X_t are $(k^2h + k)\times 1$ and $(kh + 1)\times 1$ vectors, respectively, with $h = 2\bar{p}+1$. Under the echelon form restrictions (2.8) through (2.11), the representation (2.14) implies a unique $(k^2h + k) \times r_{\bar{p}}$ full rank columns matrix R formed by $r_{\bar{p}}$-distinct selected vectors from the identity matrix of order $(k^2h + k)$ such that $R'R = I_{r_{\bar{p}}}$ and $\beta = R\eta$, where η is a $r_{\bar{p}} \times 1$ vector of free varying parameters and $r_{\bar{p}} < (k^2h + k)$. Hence (2.14) can be alternatively expressed as

$$y_t = \big[X_t' \otimes I_k\big]R\eta + u_t \tag{2.17}$$

where $\big[X_t' \otimes I_k\big]R$ is a $k \times r_{\bar{p}}$ matrix. Further, the echelon form ensures that $R'\big[X_t \otimes I_k\big]$ has a non singular covariance matrix, so that

$$\operatorname{rank}\big\{R'\big[\Gamma_X \otimes I_k\big]R\big\} = r_{\bar{p}} \tag{2.18}$$

where $\Gamma_X = \mathsf{E}\big[X_t X_t'\big]$.

Now, let $y = \big[y_1', \ldots, y_T'\big]'$, $X = \big[X_1, \ldots, X_T\big]$ and $u = \big[u_1', \ldots, u_T'\big]'$. Then the corresponding stacked form of (2.17) is

$$y = \big[X' \otimes I_k\big]R\eta + u \tag{2.19}$$

where $\big[X' \otimes I_k\big]R$ is a $(kT) \times r_{\bar{p}}$ matrix. In the sequel, we shall assume that

$$\text{rank}\big\{\big[X' \otimes I_k\big]R\big\} = r_{\bar{p}} \text{ with probability 1} \tag{2.20}$$

and under the assumption that the process is regular with continuous distribution, the latter statement must hold.

2.3. Regularity assumptions

Further assumptions on the innovation process and the truncation lag of the first step autoregression are needed in order to establish the consistency as well as the asymptotic distribution of the linear estimators defined below. For that we state the assumptions we shall consider in the sequel.

Assumption 2.1 STRONG WHITE NOISE INNOVATIONS. *The vectors u_t, $t \in \mathbb{Z}$, are independent and identically distributed (i.i.d.) with mean zero, covariance matrix Σ_u and continuous distribution.*

Assumption 2.2 UNIFORM BOUNDEDNESS OF FOURTH MOMENTS. *There is a finite constant m_4 such that, for all $1 \leq i,\ j,\ r,\ s \leq k$,*

$$\mathsf{E}\,|u_{i,t}u_{j,t}u_{r,t}u_{s,t}| \leq m_4 < \infty\,, \text{ for all } t.$$

Assumption 2.3 AUTOREGRESSIVE TRUNCATION LAG OF ORDER LESS THAN $T^{1/2}$. *n_T is a function of T such that*

$$n_T \to \infty \text{ and } n_T^2/T \to 0 \text{ as } T \to \infty \tag{2.21}$$

and, for some $c > 0$ and $0 < \delta_1 < 1/2$,

$$n_T \geq cT^{\delta_1} \text{ for } T \text{ sufficiently large.} \tag{2.22}$$

Assumption 2.4 DECAY RATE OF TRUNCATED AUTOREGRESSIVE COEFFICIENTS. *The coefficients of the autoregressive representation* (2.2) *satisfy*

$$n_T^{1/2} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T, n_T \to \infty. \tag{2.23}$$

Assumption 2.5 AUTOREGRESSIVE TRUNCATION LAG OF ORDER LESS THAN $T^{1/3}$. *n_T is a function of T such that*

$$n_T \to \infty \text{ and } n_T^3/T \to 0 \text{ as } T \to \infty \tag{2.24}$$

and, for some $c > 0$ and $0 < \delta_2 < 1/3$,

$$n_T \geq cT^{\delta_2} \text{ for } T \text{ sufficiently large.} \tag{2.25}$$

Assumption 2.6 STRONGER DECAY RATE OF TRUNCATED AUTOREGRESSIVE COEFFICIENTS. *The coefficients of the autoregressive representation* (2.2) *satisfy*

$$T^{1/2} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T, n_T \to \infty. \tag{2.26}$$

Assumption 2.7 FURTHER STRONGER DECAY RATE OF TRUNCATED AUTOREGRESSIVE COEFFICIENTS. *The coefficients of the autoregressive representation* (2.2) *satisfy*

$$T^{\delta_3} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T, n_T \to \infty \tag{2.27}$$

for some $1/2 < \delta_3 < 1$.

Assumption **2.1** means that we have a strong VARMA process, while Assumption **2.2** on moments of order four will ensure that the empirical autocovariances of the process have finite variances. Assumption **2.3** implies that n_T goes to infinity at a rate slower than $T^{1/2}$; for example, the assumption is satisfied if $n_T = cT^{\delta}$ with $0 < \delta_1 \leq \delta < 1/2$. Assumption **2.4** characterizes the rate of decay of autoregressive coefficients in relation with n_T. While Assumptions **2.5** and **2.6** are stronger ver-

sions of Assumptions **2.3** and **2.4**, respectively. Assumption **2.7** implies that for any constant $1/2 < \delta \leq \delta_3$ (with $\delta_3 < 1$) the truncated sum $T^{\delta} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\|$ converges to zero as T and n_T go to infinity.

Although the above assumptions are sufficient to show consistency of the two-step linear estimator, another assumption is needed to show the asymptotic normality of its distribution.

Assumption 2.8 AUTOREGRESSIVE TRUNCATION LAG OF ORDER LESS THAN $T^{1/4}$. *n_T is a function of T such that*

$$n_T \to \infty \text{ and } n_T^4/T \to 0 \text{ as } T \to \infty. \tag{2.28}$$

The latter assumption means that n_T goes to infinity at a rate slower than $T^{1/4}$; for example, it is satisfied if $n_T = cT^{\delta}$ with $0 < \bar{\delta} \leq \delta < 1/4$. It is easy to see that (2.28) entails (2.24) and (2.21). Finally, it is worthwhile to note that (2.23) holds for VARMA processes whenever $n_T = cT^{\delta}$ with $c > 0$ and $\delta > 0$, *i.e.*

$$T^{\delta} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T \to \infty, \quad \text{for all } \delta > 0. \tag{2.29}$$

This follows from the exponential decay of $\|\Pi_\tau\|$ for VARMA processes.

3. Generalized two-step linear estimation

In this section, we propose a generalized two-step linear regression method that yields consistent estimates of echelon-form VARMA models with known order. The Kronecker indices characterizing this parameterization are given and we focus on the estimation of the autoregressive and moving average coefficients.

Let $\left(y_{-n_T+1}, \ldots, y_T\right)$ be a random sample of size $n_T + T$ where n_T is a function of T such that n_T grows to infinity as T goes to infinity. Now, consider the "*long*"

multivariate linear autoregressive model of lag-order n_T :

$$y_t = \mu_{\Pi}(n_T) + \sum_{\tau=1}^{n_T} \Pi_\tau y_{t-\tau} + u_t(n_T) \tag{3.1}$$

where $\mu_{\Pi}(n_T) = \left(I_k - \sum_{\tau=1}^{n_T} \Pi_\tau\right)\mu_y$ and

$$u_t(n_T) = \sum_{\tau=n_T+1}^{\infty} \Pi_\tau\left(y_{t-\tau} - \mu_y\right) + u_t. \tag{3.2}$$

Setting $Y_t(n_T) = \left[1, y'_{t-1}, \ldots, y'_{t-n_T}\right]'$ and $\Pi(n_T) = \left[\mu_{\Pi}(n_T), \Pi_1, \ldots, \Pi_{n_T}\right]$, then the corresponding multivariate least squares estimator is :

$$\tilde{\Pi}(n_T) = \left[\tilde{\mu}_{\Pi}(n_T), \tilde{\Pi}_1(n_T), \ldots, \tilde{\Pi}_{n_T}(n_T)\right] = \tilde{W}_Y(n_T)\,\tilde{\Gamma}_Y(n_T)^{-1} \tag{3.3}$$

where $\tilde{W}_Y(n_T) = T^{-1}\sum_{t=1}^{T} y_t Y_t(n_T)'$ and $\tilde{\Gamma}_Y(n_T) = T^{-1}\sum_{t=1}^{T} Y_t(n_T)\,Y_t(n_T)'$. This estimation can be performed by running k separate univariate linear regressions, one for each component $y_{k,t}$. The Yule-Walker estimates of the theoretical coefficients Π_τ could also be considered. Set also $\Gamma_Y(n_T) = \mathsf{E}\left[Y_t(n_T)\,Y_t(n_T)'\right]$ and consider the norm $\|.\|_1$ such that, for any given matrix A, $\|A\|_1$ is the largest eigenvalue of $A'A$, so that $\|A\|_1 = \sup_{x\neq 0}\left\{\frac{\|Ax\|}{\|x\|}\right\}$. Then we have the following proposition.

Proposition 3.1 REGRESSOR COVARIANCE ESTIMATOR CONSISTENCY-FIRST STAGE. *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VAR representation* (3.1)*. Then, under the Assumptions* **2.1** *to* **2.3**, *we have*

$$\left\|\Gamma_Y(n_T)^{-1}\right\| = O_p(1), \tag{3.4}$$

$$\left\|\tilde{\Gamma}_Y(n_T)^{-1} - \Gamma_Y(n_T)^{-1}\right\| = O_p\left(n_T/T^{1/2}\right), \tag{3.5}$$

$$\left\|\tilde{\Gamma}_Y(n_T)^{-1} - \Gamma_Y(n_T)^{-1}\right\|_1 = O_p\left(n_T/T^{1/2}\right). \tag{3.6}$$

Proofs are supplied in Appendix A. If Assumption **2.4** is also satisfied, we can give extensions of Theorems **1** and **2.1** of Lewis and Reinsel (1985) and Paparoditis (1996), respectively, which allow for the presence of a drift.

Theorem 3.1 CONSISTENCY OF VAR COEFFICIENT ESTIMATES. *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VAR representation* (3.1)*. Then, under the Assumptions* **2.1** *to* **2.4**, *we have :*

$$\left\| \tilde{\Pi}\left(n_T\right) - \Pi\left(n_T\right) \right\| = o_p(1). \tag{3.7}$$

If, furthermore, Assumption **2.6** *holds, then*

$$\left\| \tilde{\Pi}\left(n_T\right) - \Pi\left(n_T\right) \right\| = O_p\left(n_T^{1/2}/T^{1/2}\right). \tag{3.8}$$

Now, let $l\left(n_T\right)$ be a sequence of $(k^2 n_T + k) \times 1$ vectors such that

$$0 < M_1 \leq \left\| l\left(n_T\right) \right\|^2 \leq M_2 < \infty \quad \text{for } n_T = 1, 2, \ldots \tag{3.9}$$

Set also

$$\tilde{S}_Y\left(n_T\right) = T^{1/2} l\left(n_T\right)' \operatorname{vec}\left[\tilde{\Pi}\left(n_T\right) - \Pi\left(n_T\right)\right] = T^{1/2} l\left(n_T\right)' \operatorname{vec}\left[\tilde{\Omega}_Y\left(n_T\right) \tilde{\Gamma}_Y\left(n_T\right)^{-1}\right] \tag{3.10}$$

and

$$S_Y\left(n_T\right) = T^{1/2} l\left(n_T\right)' \operatorname{vec}\left[\Omega_Y\left(n_T\right) \Gamma_Y\left(n_T\right)^{-1}\right], \tag{3.11}$$

with $\tilde{\Omega}_Y\left(n_T\right) = T^{-1} \sum_{t=1}^{T} u_t\left(n_T\right) Y_t\left(n_T\right)'$ and $\Omega_Y\left(n_T\right) = T^{-1} \sum_{t=1}^{T} u_t Y_t\left(n_T\right)'$. Then we have the following asymptotic equivalence.

Proposition 3.2 ASYMPTOTIC EQUIVALENCE-FIRST STAGE. *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VAR representation* (3.1)*. Then, under the Assumptions* **2.1**, **2.2**, **2.5** *and* **2.6**, *we have :*

$$\left\| \tilde{S}_Y\left(n_T\right) - S_Y\left(n_T\right) \right\| = o_p(1). \tag{3.12}$$

If, furthermore, Assumption **2.7** *holds, then*

$$\left\| \tilde{S}_Y\left(n_T\right) - S_Y\left(n_T\right) \right\| \doteq O_p\left(n_T^{3/2}/T^{1/2}\right). \tag{3.13}$$

Moreover, the next theorem shows that asymptotic normality holds as an immediate consequence of the proposition above. This proposition and the following theorem are generalizations of Theorems **2** and **4** of Lewis and Reinsel (1985), to the case where the drift parameter (constant) is present.

Theorem 3.2 ASYMPTOTIC DISTRIBUTION OF VAR COEFFICIENT ESTIMATES. *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VAR representation* (3.1)*. Then, under the Assumptions* **2.1**, **2.2**, **2.5**, **2.6** *and* **2.7**, *we have :*

$$\frac{T^{1/2} l(n_T)' \operatorname{vec}\left[\tilde{\Pi}(n_T) - \Pi(n_T)\right]}{\left\{l(n_T)' Q_Y(n_T) l(n_T)\right\}^{1/2}} \underset{T\to\infty}{\longrightarrow} N[0,1] \tag{3.14}$$

where

$$Q_Y(n_T) = \Gamma_Y(n_T)^{-1} \otimes \Sigma_u. \tag{3.15}$$

Note that a possible choice for the sequence n_T that satisfies both assumptions **2.5** and **2.6** is for example $n_T = T^{1/\varepsilon}$ with $\varepsilon > 3$. On the other hand $n_T = \ln \ln T$, as suggested by Hannan and Kavalieris (1984b), is not a permissible choice because in general $T^{1/2} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\|$ does not approach zero as $T \to \infty$.

Let

$$\tilde{u}_t(n_T) = y_t - \tilde{\mu}_\Pi(n_T) - \sum_{\tau=1}^{n_T} \tilde{\Pi}_\tau(n_T)\, y_{t-\tau} \tag{3.16}$$

be the estimated least-squares residuals obtained from the long autoregression (3.1), and

$$\tilde{\Sigma}_u(n_T) = \frac{1}{T} \sum_{t=1}^{T} \tilde{u}_t(n_T)\, \tilde{u}_t(n_T)' \tag{3.17}$$

the estimated covariance matrix based on these residuals. Then $\tilde{\Sigma}_u(n_T)$ satisfies the following convergence properties.

Proposition 3.3 INNOVATION COVARIANCE ESTIMATOR CONSISTENCY-FIRST STAGE. *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11)*. Then, under*

the assumptions **2.1** *to* **2.4**, *we have :*

$$\left\|\tilde{\Sigma}_u\left(n_T\right)-\Sigma_u\right\|=\left\|\tilde{\Sigma}_u\left(n_T\right)^{-1}-\Sigma_u^{-1}\right\|=O_p\left(n_T/T^{1/2}\right). \tag{3.18}$$

The asymptotic equivalence stated in the above proposition suggests that we may be able to consistently estimate the parameters of the VARMA model in (2.14) by replacing the unobserved lagged innovations $u_{t-1}, \ldots, u_{t-\bar{p}}$ with their corresponding first stage estimates $\tilde{u}_{t-1}\left(n_T\right), \ldots, \tilde{u}_{t-\bar{p}}\left(n_T\right)$. So, in order to estimate the coefficients Φ_i and Θ_j of the VARMA process, (2.14) can be rewritten as

$$y_t=\mu_\Phi+\left(I_k-\Phi_0\right)\tilde{v}_t\left(n_T\right)+\sum_{i=1}^{\bar{p}}\Phi_i y_{t-i}+\sum_{j=1}^{\bar{p}}\Theta_j\tilde{u}_{t-j}\left(n_T\right)+e_t\left(n_T\right) \tag{3.19}$$

or equivalently,

$$y_t=\left[\tilde{X}_t\left(n_T\right)'\otimes I_k\right]R\eta+e_t\left(n_T\right) \tag{3.20}$$

where

$$\begin{aligned}
\tilde{v}_t\left(n_T\right) &= y_t-\tilde{u}_t\left(n_T\right), && (3.21)\\
e_t\left(n_T\right) &= \tilde{u}_t\left(n_T\right)+\sum_{j=0}^{\bar{p}}\Theta_j\left[u_{t-j}-\tilde{u}_{t-j}\left(n_T\right)\right], && (3.22)\\
\tilde{X}_t\left(n_T\right) &= \left[1,\tilde{v}_t\left(n_T\right)',y_{t-1}',\ldots,y_{t-\bar{p}}',\tilde{u}_{t-1}\left(n_T\right)',\ldots,\tilde{u}_{t-\bar{p}}\left(n_T\right)'\right]'. && (3.23)
\end{aligned}$$

This yields as a second step estimator of η the generalized least squares (GLS) estimate, obtained by solving

$$\tilde{\eta}=\arg\min_{\eta}\sum_{t=1}^{T}e_t\left(n_T\right)\tilde{\Sigma}_u\left(n_T\right)^{-1}e_t\left(n_T\right)'. \tag{3.24}$$

$\tilde{\eta}$ has the form :

$$\tilde{\eta}=\tilde{Q}_X\left(n_T\right)\tilde{W}_X\left(n_T\right) \tag{3.25}$$

where

$$\tilde{Q}_X(n_T) = \left\{R'\tilde{\Upsilon}_X(n_T)R\right\}^{-1}, \tag{3.26}$$

$$\tilde{\Upsilon}_X(n_T) = \tilde{\Gamma}_X(n_T) \otimes \tilde{\Sigma}_u(n_T)^{-1}, \tag{3.27}$$

$$\tilde{\Gamma}_X(n_T) = T^{-1}\sum_{t=1}^{T} \tilde{X}_t(n_T)\tilde{X}_t(n_T)', \tag{3.28}$$

$$\tilde{W}_X(n_T) = T^{-1}\sum_{t=1}^{T} R'\left[\tilde{X}_t(n_T) \otimes I_k\right]\tilde{\Sigma}_u(n_T)^{-1} y_t. \tag{3.29}$$

Setting

$$\tilde{\Omega}_X(n_T) = T^{-1}\sum_{t=1}^{T} R'\left[\tilde{X}_t(n_T) \otimes I_k\right]\tilde{\Sigma}_u(n_T)^{-1} e_t(n_T), \tag{3.30}$$

one sees easily that

$$\tilde{\eta} - \eta = \tilde{Q}_X(n_T)\tilde{\Omega}_X(n_T). \tag{3.31}$$

On using the inequality $\|AB\|^2 \leq \|A\|_1^2 \|B\|^2$, where A and B are any two conformable matrices, we get

$$\|\tilde{\eta} - \eta\| \leq \left\|\tilde{Q}_X(n_T)\right\|_1 \left\|\tilde{\Omega}_X(n_T)\right\|. \tag{3.32}$$

Now, define

$$\Upsilon_X = \Gamma_X \otimes \Sigma_u^{-1}, \quad Q_X = \left\{R'\Upsilon_X R\right\}^{-1}, \tag{3.33}$$

$$\Omega_X = T^{-1}\sum_{t=1}^{T} R'\left[X_t \otimes I_k\right]\Sigma_u^{-1} u_t. \tag{3.34}$$

Obviously, by the regularity assumption Q_X^{-1} is positive definite, and to study the convergence and distributional properties of $(\tilde{\eta} - \eta)$ we need first to establish the following proposition.

Proposition 3.4 REGRESSOR COVARIANCE ESTIMATOR CONSISTENCY-SECOND STAGE. *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *Then, under*

the assumptions **2.1** *to* **2.4**, *we have :*

$$\left\|\tilde{Q}_X\left(n_T\right) - Q_X\right\|_1 = O_p\left(n_T/T^{1/2}\right). \tag{3.35}$$

The latter proposition shows that the regressor matrix $\tilde{X}_t\left(n_T\right)$ as well as the covariance matrix $\tilde{Q}_X\ \left(n_T\right)$ – based on approximate innovations (estimated from a long autoregression) – are all asymptotically equivalent to their corresponding analogous based on true innovations, according to the rate $n_T/T^{1/2}$. This suggests that $\tilde{\eta}$ converges to η. The following theorem gives the appropriate rate of such convergence.

Theorem 3.3 CONSISTENCY OF TWO-STEP ESTIMATORS. *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *Then, under the assumptions* **2.1** *to* **2.8**, *we have :*

$$\left\|\tilde{\eta} - \eta\right\| = O_p\left(T^{-1/2}\right). \tag{3.36}$$

In order to derive the asymptotic distribution for $\tilde{\eta}$, we shall first establish the asymptotic equivalence between the following random vectors

$$\tilde{S}_X\left(n_T\right) = T^{1/2}\tilde{Q}_X\left(n_T\right)\tilde{\Omega}_X\left(n_T\right)\ , \quad S_X = T^{1/2}Q_X\Omega_X. \tag{3.37}$$

Proposition 3.5 ASYMPTOTIC EQUIVALENCE-SECOND STAGE. *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *Then, under the assumptions* **2.1** *to* **2.8**, *we have :*

$$\left\|\tilde{S}_X\left(n_T\right) - S_X\right\| = O_p\left(n_T^2/T^{1/2}\right). \tag{3.38}$$

Finally, we provide the asymptotic distribution of the two-step generalized estimators in the following theorem.

Theorem 3.4 ASYMPTOTIC DISTRIBUTION OF TWO-STEP ESTIMATORS. *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *If the assumptions* **2.1**

to **2.8** *are satisfied, then*

$$T^{1/2}(\tilde{\eta}-\eta) \underset{T\to\infty}{\longrightarrow} N[0, Q_X], \quad T^{1/2}(\tilde{\beta}-\beta) \underset{T\to\infty}{\longrightarrow} N[0, V_X] \tag{3.39}$$

where $\tilde{\beta} = R\tilde{\eta}$ *and* $V_X = RQ_XR'$.

Now set

$$\tilde{\Sigma}_e(n_T) = \frac{1}{T}\sum_{t=1}^{T} \tilde{e}_t(n_T)\,\tilde{e}_t(n_T)' \tag{3.40}$$

where

$$\tilde{e}_t(n_T) = y_t - \left[\tilde{X}_t(n_T)' \otimes I_k\right] R\tilde{\eta}. \tag{3.41}$$

Then the following proposition ensures that $\tilde{\Sigma}_e(n_T)$ and $\tilde{\Sigma}_e(n_T)^{-1}$ are consistent estimators of Σ_u and Σ_u^{-1}, respectively.

Proposition 3.6 INNOVATION COVARIANCE ESTIMATOR CONSISTENCY-SECOND STAGE. *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *Then, under the assumptions* **2.1** *to* **2.8**, *we have :*

$$\left\|\tilde{\Sigma}_e(n_T) - \Sigma_u\right\| = \left\|\tilde{\Sigma}_e(n_T)^{-1} - \Sigma_u^{-1}\right\| = O_p\left(n_T/T^{1/2}\right). \tag{3.42}$$

4. Asymptotic efficiency

The two-stage linear estimator derived above is not efficient under Gaussian innovations. To allow for asymptotic efficiency [as in the fourth-stage of Hannan and Kavalieris (1984b)], we shall perform a third-stage estimation by considering a linear regression that we shall describe in the present section.

Unlike Hannan and Kavalieris (1984b) who did not provide an explicit justification of the suggested fourth-stage estimators, we supply here a derivation of such estimators. This will be useful to show the asymptotic efficiency of these estimates for Gaussian errors. In line with their procedure, at the specification level, the fourth stage estimation they suggested to achieve asymptotic efficiency is complicated and

heavy to implement even in small systems. The final form of these estimators does not show the echelon form zero-restrictions. Further, the estimated covariance matrix of these estimators has a complicated form which is difficult to compute in practice.

In contrast, we provide simple, compact, and efficient echelon-form estimators which can be easily computed by running a simple linear regression. Hence, one might consider further linear regressions as they are costless. Furthermore, we provide a simple formula to estimate its covariance matrix.

Recall that our initial problem consists in minimizing an objective function which is nonlinear in the parameters η : we have to solve

$$\min_{\eta} \sum_{t=1}^{T} u_t' \Sigma_u^{-1} u_t \tag{4.1}$$

where $u_t = \sum_{\tau=0}^{\infty} \Lambda_\tau(\eta) \left[\Phi_0 \left(y_{t-\tau} - \mu_y\right) - \sum_{i=1}^{\bar{p}} \Phi_i \left(y_{t-i-\tau} - \mu_y\right)\right]$. Setting

$$u_t(\eta) = \sum_{\tau=0}^{t-1} \Lambda_\tau(\eta) \left[\Phi_0\big(y_{t-\tau} - \mu_y\big) - \sum_{i=1}^{\bar{p}} \Phi_i\big(y_{t-i-\tau} - \mu_y\big)\right], \tag{4.2}$$

one sees easily that $\|u_t - u_t(\eta)\| = O_p(\rho^t)$, since it can be shown that

$$\mathsf{E}\|u_t - u_t(\eta)\| \leq \sum_{\tau=t}^{\infty} \|\Lambda_\tau(\eta)\| \, \|\Phi(\bar{p})\| \, \mathsf{E}\left\|Y_{t-\tau}^a(\bar{p})\right\| = O\left(\rho^t\right) \tag{4.3}$$

where $\Phi(\bar{p}) = [\Phi_0, -\Phi_1, \ldots, -\Phi_{\bar{p}}]$, $Y_t^a(\bar{p}) = \left[y_t^{a\prime}, y_{t-1}^{a\prime}, \ldots, y_{t-\bar{p}}^{a\prime}\right]'$ with $y_t^a = \left(y_t - \mu_y\right)$; see the proof of Theorem **3.1**. This suggests replacing the problem (4.1) by

$$\min_{\eta} \sum_{t=1}^{T} u_t(\eta)' \, \Sigma_u^{-1} u_t(\eta). \tag{4.4}$$

Note also that (3.41) can be alternatively expressed, as in (3.22), as

$$\tilde{e}_t(n_T) = \tilde{u}_t(n_T) + \sum_{j=0}^{\bar{p}} \tilde{\Theta}_j \left[\tilde{u}_{t-j} - \tilde{u}_{t-j}(n_T)\right], \tag{4.5}$$

so that, using the two-step estimate $\tilde{\eta}$, the estimated model

$$y_t = \left[\tilde{X}_t\left(n_T\right)' \otimes I_k\right] R\tilde{\eta} + \tilde{e}_t\left(n_T\right) \tag{4.6}$$

takes the form

$$y_t = \tilde{\mu}_{\Phi} + \left(I_k - \tilde{\Phi}_0\right)\tilde{v}_t + \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i y_{t-i} + \sum_{j=1}^{\bar{p}} \tilde{\Theta}_j \tilde{u}_{t-j} + \tilde{u}_t \tag{4.7}$$

where $\tilde{v}_t = y_t - \tilde{u}_t$ with

$$\tilde{u}_t = \sum_{\tau=0}^{\infty} \Lambda_\tau\left(\tilde{\eta}\right)\left[\tilde{\Phi}_0\left(y_{t-\tau} - \tilde{\mu}_y\right) - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i\left(y_{t-i-\tau} - \tilde{\mu}_y\right)\right], \tag{4.8}$$

$\tilde{\mu}_y = \tilde{\Phi}\left(1\right)^{-1}\tilde{\mu}_{\Phi}$, $\tilde{\Phi}\left(1\right) = \tilde{\Phi}_0 - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i$ and $\sum_{j=0}^{\infty} \Lambda_\tau\left(\tilde{\eta}\right) z^\tau = \tilde{\Theta}\left(z\right)^{-1}$, where $\tilde{\mu}_{\Phi}$, $\tilde{\Phi}_i$ and $\tilde{\Theta}_j$ stand for the second-stage estimates of μ_{Φ}, Φ_i and Θ_j, respectively. In view of (4.7) and (4.8), it seems obvious that the second step estimator $\tilde{\eta}$ may be used as initial value in the minimization algorithm when seeking for the nonlinear generalized least squares estimates. As for u_t and $u_t\left(\eta\right)$, we can approximate $\tilde{u}_t$ with

$$\begin{aligned} u_t\left(\tilde{\eta}\right) &= \sum_{\tau=0}^{t-1} \Lambda_\tau\left(\tilde{\eta}\right)\left[\tilde{\Phi}_0 y_{t-\tau} - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i y_{t-i-\tau} - \tilde{\mu}_{\Phi}\right] \\ &= \sum_{\tau=0}^{t-1} \Lambda_\tau\left(\tilde{\eta}\right)\left[\tilde{\Phi}_0\left(y_{t-\tau} - \tilde{\mu}_y\right) - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i\left(y_{t-i-\tau} - \tilde{\mu}_y\right)\right]. \end{aligned} \tag{4.9}$$

This also can be either determined recursively using

$$\begin{aligned} u_t\left(\tilde{\eta}\right) &= \tilde{\Phi}_0^{-1}\tilde{e}_t\left(n_T\right) + \left(I_k - \tilde{\Phi}_0^{-1}\right)\tilde{u}_t\left(n_T\right) + \sum_{j=1}^{\bar{p}} \tilde{\Phi}_0^{-1}\tilde{\Theta}_j\left[\tilde{u}_{t-j}\left(n_T\right) - u_{t-j}\left(\tilde{\eta}\right)\right] \\ &= y_t - \tilde{\Phi}_0^{-1}\left[\tilde{\mu}_{\Phi} + \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i y_{t-i} + \sum_{j=1}^{\bar{p}} \tilde{\Theta}_j u_{t-j}\left(\tilde{\eta}\right)\right], \end{aligned} \tag{4.10}$$

with initial values $u_t(\tilde{\eta}) = 0$, $t \leq \bar{p}$. Now, set

$$\tilde{\Sigma}_u(\tilde{\eta}) = \frac{1}{T}\sum_{t=1}^{T} u_t(\tilde{\eta})\, u_t(\tilde{\eta})'. \tag{4.11}$$

In order to establish the rate of convergence of the above innovation covariance matrix estimator to the true one, we will use the following lemma.

Lemma 4.1 CONSISTENCY OF NONLINEAR FUNCTION. *Let $\breve{\eta}$ be a $\sqrt{T}$-consistent estimator for η, i.e.*

$$T^{1/2}\|\breve{\eta} - \eta\| = O_p(1),$$

where $\|.\|$ denotes the Schur norm. Then there exists a real constant $\kappa > 0$ such that

$$T^{1/2}(1+\kappa^{-1})^{\tau}\|\Lambda_\tau(\breve{\eta}) - \Lambda_\tau(\eta)\| = O_p(1), \ \forall \tau \in \mathbb{Z}. \tag{4.12}$$

Using the above lemma, we get the following convergence results for the estimators of Σ_u and Σ_u^{-1}.

Proposition 4.1 FILTERED INNOVATION COVARIANCE ESTIMATOR CONSISTENCY-SECOND STAGE. *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *Then, under the assumptions* **2.1** *to* **2.8**, *we have :*

$$\left\|\tilde{\Sigma}_u(\tilde{\eta}) - \Sigma_u\right\| = \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1}\right\| = O_p\big(T^{-1/2}\big). \tag{4.13}$$

Now, consider the following lemma.

Lemma 4.2 RELATIONAL IDENTITY BETWEEN FILTERED RESIDUALS. *Let η^0 and η^1 be two distinct values of η. Then*

$$u_t(\eta^1) - u_t(\eta^0) = -Z_t^{\circ}(\eta^1, \eta^0)'\big(\eta^1 - \eta^0\big) \tag{4.14}$$

where

$$Z_t^{\circ}(\eta^1,\eta^0)=\sum_{\tau=0}^{t-1}R'\left[X_{t-\tau}(\eta^1)\otimes\Lambda_\tau(\eta^0)'\right], \tag{4.15}$$

$X_t(\eta^1)=\left[1,v_t(\eta^1)',y'_{t-1},\ldots,y'_{t-\bar{p}},u_{t-1}(\eta^1)',\ldots,u_{t-\bar{p}}(\eta^1)'\right]'$ *and* $v_t(\eta^1)=y_t-u_t(\eta^1)$.

By the latter lemma, we can write :

$$u_t\left(\tilde{\eta}\right)-u_t\left(\eta\right)=-Z_t^{\circ}\left(\tilde{\eta},\eta\right)'\left(\tilde{\eta}-\eta\right) \tag{4.16}$$

where

$$Z_t^{\circ}\left(\tilde{\eta},\eta\right)=\sum_{\tau=0}^{t-1}R'\left[X_{t-\tau}\left(\tilde{\eta}\right)\otimes\Lambda_\tau\left(\eta\right)'\right] \tag{4.17}$$

with $X_t\left(\tilde{\eta}\right)=\left[1,v_t\left(\tilde{\eta}\right)',y'_{t-1},\ldots,y'_{t-\bar{p}},u_{t-1}\left(\tilde{\eta}\right)',\ldots,u_{t-\bar{p}}\left(\tilde{\eta}\right)'\right]'$ and $v_t\left(\tilde{\eta}\right)=y_t-u_t\left(\tilde{\eta}\right)$. Hence (4.16) can be easily rearranged to obtain in a linear regression form

$$w_t\left(\tilde{\eta}\right)=Z_t\left(\tilde{\eta}\right)'\eta+z_t\left(\tilde{\eta},\eta\right) \tag{4.18}$$

where

$$\begin{aligned} w_t\left(\tilde{\eta}\right) &= u_t\left(\tilde{\eta}\right)+Z_t\left(\tilde{\eta}\right)'\tilde{\eta}, && (4.19)\\ Z_t\left(\tilde{\eta}\right) &= \sum_{\tau=0}^{t-1}R'\left[X_{t-\tau}\left(\tilde{\eta}\right)\otimes\Lambda_\tau\left(\tilde{\eta}\right)'\right], && (4.20)\\ z_t\left(\tilde{\eta},\eta\right) &= u_t\left(\eta\right)+\left[Z_t\left(\tilde{\eta}\right)-Z_t^{\circ}\left(\tilde{\eta},\eta\right)\right]'\left(\tilde{\eta}-\eta\right). && (4.21) \end{aligned}$$

This suggests considering a third-stage estimator, namely the multivariate GLS estimator of η obtained by regressing $\tilde{\Sigma}_u\left(\tilde{\eta}\right)^{-1/2}w_t\left(\tilde{\eta}\right)$ on $\tilde{\Sigma}_u\left(\tilde{\eta}\right)^{-1/2}Z_t\left(\tilde{\eta}\right)'$:

$$\hat{\eta}=\tilde{Q}_X\left(\tilde{\eta}\right)\tilde{W}_X\left(\tilde{\eta}\right) \tag{4.22}$$

where

$$\tilde{Q}_X\left(\tilde{\eta}\right)=\left\{T^{-1}\sum_{t=1}^{T}Z_t\left(\tilde{\eta}\right)\tilde{\Sigma}_u\left(\tilde{\eta}\right)^{-1}Z_t\left(\tilde{\eta}\right)'\right\}^{-1}, \tag{4.23}$$

$$\tilde{W}_X\ (\tilde{\eta}) \ = \ T^{-1}\sum_{t=1}^{T} Z_t\,(\tilde{\eta})\,\tilde{\Sigma}_u\,(\tilde{\eta})^{-1}\,w_t\,(\tilde{\eta})\,. \tag{4.24}$$

In view of (4.19), we can also write

$$\hat{\eta} = \tilde{\eta} + \tilde{Q}_X\,(\tilde{\eta})\,\tilde{\Omega}_X\ (\tilde{\eta}) \tag{4.25}$$

where

$$\tilde{\Omega}_X\ (\tilde{\eta}) = T^{-1}\sum_{t=1}^{T} Z_t\,(\tilde{\eta})\,\tilde{\Sigma}_u\,(\tilde{\eta})^{-1}\,u_t\,(\tilde{\eta})\,. \tag{4.26}$$

It is worth noting that, under Gaussian errors, $\hat{\eta}$ is asymptotically equivalent to ML estimator, for $\frac{\partial u_t(\tilde{\eta})}{\partial \eta'} = -Z_t\,(\tilde{\eta})'$; see (4.19). In view of (4.25), $\hat{\eta}$ corresponds to the estimator based on the scoring method after one iteration.

Now, let

$$\tilde{Q}_X^{\circ}\,(\tilde{\eta}) \ = \ \left\{T^{-1}\sum_{t=1}^{T} Z_t^{\circ}\,(\tilde{\eta},\eta)\,\tilde{\Sigma}_u\,(\tilde{\eta})^{-1}\,Z_t^{\circ}\,(\tilde{\eta},\eta)'\right\}^{-1}, \tag{4.27}$$

$$\tilde{\Omega}_X^{\circ}\,(\tilde{\eta}) \ = \ T^{-1}\sum_{t=1}^{T} Z_t^{\circ}\,(\tilde{\eta},\eta)\,\tilde{\Sigma}_u\,(\tilde{\eta})^{-1}\,u_t\,(\tilde{\eta})\,, \tag{4.28}$$

$$\tilde{\Omega}_X^{\bullet}\,(\tilde{\eta}) \ = \ T^{-1}\sum_{t=1}^{T} Z_t^{\circ}\,(\tilde{\eta},\eta)\,\tilde{\Sigma}_u\,(\tilde{\eta})^{-1}\,u_t\,(\eta)\,, \tag{4.29}$$

$$Q_X\,(\eta) = \left\{\mathsf{E}\left[Z_t\,(\eta)\,\Sigma_u^{-1}Z_t\,(\eta)'\right]\right\}^{-1}, \quad \Omega_X\,(\eta) = T^{-1}\sum_{t=1}^{T} Z_t\,(\eta)\,\Sigma_u^{-1}u_t, \tag{4.30}$$

$$Z_t\,(\eta) = \sum_{\tau=0}^{\infty} R'\left[X_{t-\tau}\otimes \Lambda_\tau\,(\eta)'\right]. \tag{4.31}$$

Using Lemma **4.2**, equation (4.25) can be rewritten as

$$\hat{\eta} - \eta = \tilde{Q}_X\,(\tilde{\eta})\,\tilde{\Omega}_X\,(\tilde{\eta}) + \tilde{Q}_X^{\circ}\,(\tilde{\eta})\left[\tilde{\Omega}_X^{\bullet}\,(\tilde{\eta}) - \tilde{\Omega}_X^{\circ}\,(\tilde{\eta})\right]. \tag{4.32}$$

Further, $Q_X\,(\eta)$ can be expressed as

$$Q_X\,(\eta) = \left\{R'\Upsilon_X\,(\eta)\,R\right\}^{-1} \tag{4.33}$$

where

$$\Upsilon_X(\eta) = \sum_{\tau_1=0}^{\infty}\sum_{\tau_2=0}^{\infty}\Big[\Gamma_X(\tau_1-\tau_2)\otimes A_{\tau_1}(\eta)'\,\Sigma_u^{-1}A_{\tau_2}(\eta)\Big], \tag{4.34}$$

with $\Gamma_X(\tau_1-\tau_2) = \mathsf{E}\big[X_{t-\tau_1}X_{t-\tau_2}'\big]$. By construction $Q_X(\eta)^{-1}$ is positive definite, and to study the convergence and distributional properties of $\hat{\eta}-\eta$, we first establish the following asymptotic equivalences.

Proposition 4.2 REGRESSOR COVARIANCE ESTIMATOR CONSISTENCY-THIRD STAGE. *Let $\{y_t : t\in\mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *Then, under the assumptions* **2.1** *to* **2.8**, *we have :*

$$\big\|\tilde{Q}_X(\tilde{\eta}) - \tilde{Q}_X^{\circ}(\tilde{\eta})\big\|_1 = \big\|\tilde{Q}_X^{\circ}(\tilde{\eta}) - Q_X(\eta)\big\|_1 = O_p\big(T^{-1/2}\big). \tag{4.35}$$

Then, we can give the rate of convergence of the third-stage estimator $\hat{\eta}$.

Theorem 4.1 CONSISTENCY OF THREE-STEP ESTIMATORS. *Let $\{y_t : t\in\mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *Then, under the assumptions* **2.1** *to* **2.8**, *we have :*

$$\|\hat{\eta}-\eta\| = O_p\big(T^{-1/2}\big). \tag{4.36}$$

Now, set

$$\tilde{S}_X(\tilde{\eta}) = T^{1/2}\Big\{\tilde{Q}_X(\tilde{\eta})\,\tilde{\Omega}_X(\tilde{\eta}) + \tilde{Q}_X^{\circ}(\tilde{\eta})\Big[\tilde{\Omega}_X^{\bullet}(\tilde{\eta}) - \tilde{\Omega}_X^{\circ}(\tilde{\eta})\Big]\Big\}, \tag{4.37}$$

$$S_X(\eta) = T^{1/2}Q_X(\eta)\,\Omega_X(\eta). \tag{4.38}$$

These two vectors satisfy the following asymptotic equivalence.

Proposition 4.3 ASYMPTOTIC EQUIVALENCE-THIRD STAGE. *Let $\{y_t : t\in\mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *Then, under the assumptions* **2.1** *to* **2.8**, *we*

have :

$$\left\|\tilde{S}_X(\tilde{\eta}) - S_X(\eta)\right\| = O_p\left(T^{-1/2}\right). \tag{4.39}$$

We can now establish the asymptotic normality of the third-stage estimator.

Theorem 4.2 ASYMPTOTIC DISTRIBUTION OF THREE-STEP ESTIMATORS. *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *Then, under the assumptions* **2.1** *to* **2.8**, *we have :*

$$T^{1/2}\left(\hat{\eta} - \eta\right) \underset{T\to\infty}{\longrightarrow} N\left[0, Q_X(\eta)\right], \quad T^{1/2}\left(\hat{\beta} - \beta\right) \underset{T\to\infty}{\longrightarrow} N\left[0, V_X(\eta)\right] \tag{4.40}$$

where $\hat{\beta} = R\hat{\eta}$ *and* $V_X(\eta) = RQ_X(\eta)R'$.

Again, the third-stage residuals $u_t(\hat{\eta})$ can be either recursively filtered using

$$u_t(\hat{\eta}) = z_t(\tilde{\eta}, \hat{\eta}) - \left[Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta}, \hat{\eta})\right]'(\tilde{\eta} - \hat{\eta}), \tag{4.41}$$

$$u_t(\hat{\eta}) = \sum_{\tau=0}^{t-1} \Lambda_\tau(\hat{\eta}) \left[\hat{\Phi}_0 y_{t-\tau} - \sum_{i=1}^{\bar{p}} \hat{\Phi}_i y_{t-i-\tau} - \hat{\mu}_\Phi\right], \tag{4.42}$$

or

$$u_t(\hat{\eta}) = y_t - \hat{\Phi}_0^{-1}\left[\hat{\mu}_\Phi + \sum_{i=1}^{\bar{p}} \hat{\Phi}_i y_{t-i} + \sum_{j=1}^{\bar{p}} \hat{\Theta}_j u_{t-j}(\hat{\eta})\right], \tag{4.43}$$

initiating with $u_t(\hat{\eta}) = 0$, $t \leq \bar{p}$, so that they verify

$$\hat{\Phi}(L)\, y_t = \hat{\mu}_\Phi + \hat{\Theta}(L)\, u_t(\hat{\eta}), \quad t = 1, \ldots, T, \tag{4.44}$$

or equivalently

$$\hat{\Phi}(L)\left(y_{t-\tau} - \hat{\mu}_y\right) = \hat{\Theta}(L)\, u_t(\hat{\eta}), \quad t = 1, \ldots, T \tag{4.45}$$

where $\hat{\mu}_y = \hat{\Phi}(1)^{-1}\hat{\mu}_\Phi$ and $\hat{\Phi}(1) = \hat{\Phi}_0 - \sum_{i=1}^{\bar{p}} \hat{\Phi}_i$. Therefore, the third-stage residual

covariance matrix estimator is such that

$$\tilde{\Sigma}_u(\hat{\eta}) = \frac{1}{T}\sum_{t=1}^{T} u_t(\hat{\eta})\, u_t(\hat{\eta})' \tag{4.46}$$

and its rate of convergence to the true one is finally given in the following proposition whose proof is similar to that of Proposition **4.1**.

Proposition 4.4 INNOVATION COVARIANCE ESTIMATOR CONSISTENCY-THIRD STAGE. *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *Then, under the assumptions* **2.1** *to* **2.8**, *we have :*

$$\left\| \tilde{\Sigma}_u(\hat{\eta}) - \Sigma_u \right\| = O_p\left(T^{-1/2}\right). \tag{4.47}$$

5. Model specification

So far, the dynamic indices characterizing the echelon-form representation of VARMA models are considered as given. Which is not the case in practice since they must be estimated. Several model specification procedures have been proposed to this end. However, we shall focus on information-type criteria techniques, since they are easy to implement compared to other existing methods, even in high dimensional systems. For a review of other specification procedures, see Lütkepohl and Poskitt (1996).

Model selection criteria, such as the AIC criterion [Akaike (1969, 1973, 1974b, 1977)], the BIC criterion [Schwarz (1978), Rissanen (1978)] and the HQ criterion [Hannan and Quinn (1979), Quinn (1980)] have been proposed to estimate the Kronecker indices characterizing the dynamic structure of VARMA models. These information (Akaike-type) model selection criteria, typically have the form :

$$Cr(p) = \ln\Big(\det\left\{\tilde{\Sigma}_u(p)\right\}\Big) + f(p)\frac{C(T)}{T} \tag{5.1}$$

where p is a vector of orders specifying the lag-structure of the model, $f(p)$ is the total number of the coefficient parameters implied by p, $\tilde{\Sigma}_u(p)$ is the corresponding estimated residual covariance matrix, and $C(T)$ is a function of the sample size T. Basically, such criteria incorporate penalties for the number of fitted coefficients, and they are minimized over the space of the echelon-form VARMA parameters. However, these methods may be time consuming when full-search strategies over a prespecified set of Kronecker indices are considered in big systems. Practical problems may also arise in small systems, especially, when fitting overparameterized models. Further, $C(T)$ is usually a *logarithmic* function of T, which causes the adjustment penalty or the ratio $C(T)/T$ to vanish quickly toward zero as T grows to infinity, and consequently, leads to overparameterized models, especially in highly persistent processes.

For stationary invertible multivariate ARMAX models with the echelon form representation, Hannan and Kavalieris (1984b) have proposed shortcuts to estimate the Kronecker indices within a four-step linear estimation procedure [see also Hannan and Deistler (1988)]. These shortcuts (which appear in their second and third steps) aim at identifying the Kronecker indices using prespecified upper bounds. They propose searching over several echelon-forms with identical individual indices in a first step, then varying the individual Kronecker indices in a second step, taking as upper bounds the selected first-step indices. Although the shortcuts procedure they have provided to estimate these indices dramatically reduces the number of models to be compared at the specification stages, it misses singularity problems related to overparameterized model fitting. These problems have been partially solved, recently, by Poskitt (1992) and Poskitt and Lütkepohl (1995) who assessed the Kronecker indices sequentially from smallest to largest using single-equation least squares (LS) estimation. An extension to cointegrated VARMA analysis was also given by Lütkepohl and Claessen (1997) and Lütkepohl (1999). However, methods which minimize information criteria over single equations, separately, are less informative since they tend to ignore the covariance structure of residuals implied by the full echelon form model. This yields imprecise estimates of the true Kronecker indices, hence highly inefficient

echelon-form VARMA parameter estimates.

In this section, we propose model selection criteria and shortcuts that overcome these difficulties to estimate the true Kronecker indices of echelon-form VARMA models. These information criteria are of type (5.1), and are suggested only for the first two stages of our procedure, since the third one deals with the asymptotic efficiency. As we will see below the adjustment penalties considered in these criteria are such that $C(T)$ is a function of the long-autoregression lag n_T, then an exponential function of T so that the ratio $C(T)/T$ does not vanish too quickly as T goes to infinity. This increases the probability of estimating the true Kronecker indices and eliminates possibly overparameterized models.

Now, let A and B be any two $k \times k$ positive definite symmetric matrices. These can be diagonalized as $A = P_A \Lambda_A P_A'$ and $B = P_B \Lambda_B P_B'$, where P_A and P_B are the orthogonal matrices formed by the corresponding eigenvectors of A and B,

$$\Lambda_A = \operatorname{diag}\Big\{\lambda_1(A), \ldots, \lambda_k(A)\Big\}, \quad \Lambda_B = \operatorname{diag}\Big\{\lambda_1(B), \ldots, \lambda_k(B)\Big\}, \tag{5.2}$$

and $\lambda_i(A)$ and $\lambda_i(B)$, $i = 1, \ldots, k$, are the eigenvalues (all positive) of A and B, respectively. Setting

$$x_i = \frac{\lambda_i(A) - \lambda_i(B)}{\lambda_i(B)}, \tag{5.3}$$

then

$$x_i > -1 \quad \text{for any} \quad \lambda_i(A); \lambda_i(B) > 0, \quad i = 1, \ldots, k. \tag{5.4}$$

Further, since $\ln(1 + x_i) \leq x_i$ for $x_i > -1$, one easily sees that

$$\begin{aligned}
\ln\Big(\det\{A\}\Big) &= \ln\Big(\det\{B\}\Big) + \sum_{i=1}^{k} \ln\left(1 + \frac{\lambda_i(A) - \lambda_i(B)}{\lambda_i(B)}\right) \\
&\leq \ln\Big(\det\{B\}\Big) + \sum_{i=1}^{k} \left[\frac{\lambda_i(A) - \lambda_i(B)}{\lambda_i(B)}\right] \\
&\leq \ln\Big(\det\{B\}\Big) + \left\{\sum_{i=1}^{k} \Big[\lambda_i(A) - \lambda_i(B)\Big]^2\right\}^{1/2} \left\{\sum_{i=1}^{k} \lambda_i^{-2}(B)\right\}^{1/2}
\end{aligned}$$

$$= \ln\Big(\det\{B\}\Big) + \Big\{\sum_{i=1}^{k}\Big[\lambda_i(A) - \lambda_i(B)\Big]^2\Big\}^{1/2}\Big\{\sum_{i=1}^{k}\lambda_i^2\left(B^{-1}\right)\Big\}^{1/2}, \tag{5.5}$$

hence

$$\ln\Big(\det\{A\}\Big) \leq \ln\Big(\det\{B\}\Big) + \Big\|\Lambda_A - \Lambda_B\Big\|\Big\|\Lambda_{B^{-1}}\Big\|. \tag{5.6}$$

What we learn from the last inequality is that for fixed values of $\lambda_i(A)$'s, leaving free the values of the $\lambda_i(B)$'s, the inequality is fulfilled or minimized when $\lambda_i(B) = \lambda_i(A)$, $i = 1, \ldots, k$. In other words, taking the $\lambda_i(A)$'s as constants and the $\lambda_i(B)$'s as variables, to get close and closer to $\ln(\det\{A\})$ reduces to minimize the right hand side of the last inequality with respect to the $\lambda_i(B)$'s.

Now, let Λ_{Σ_u}, $\Lambda_{\tilde{\Sigma}_u(n_T)}$ and $\Lambda_{\tilde{\Sigma}_e(n_T)}$ be the eigenvalue matrices associated with the true innovation covariance matrix, the first and the second stages estimators of the residual covariance matrices. Then using (5.6), we get

$$\ln\Big(\det\{\Sigma_u\}\Big) \leq \ln\Big(\det\{\tilde{\Sigma}_u(n_T)\}\Big) + \Big\|\Lambda_{\Sigma_u} - \Lambda_{\tilde{\Sigma}_u(n_T)}\Big\|\Big\|\Lambda_{\tilde{\Sigma}_u(n_T)^{-1}}\Big\| \tag{5.7}$$

and

$$\ln\Big(\det\{\Sigma_u\}\Big) \leq \ln\Big(\det\{\tilde{\Sigma}_e(n_T)\}\Big) + \Big\|\Lambda_{\Sigma_u} - \Lambda_{\tilde{\Sigma}_e(n_T)}\Big\|\Big\|\Lambda_{\tilde{\Sigma}_e(n_T)^{-1}}\Big\|, \tag{5.8}$$

for the first and second stages, respectively. Recall that Σ_u is the innovation covariance matrix which is a function of the parameter values of the true model. Hence, in the above two inequalities, $\ln(\det\{\Sigma_u\})$ is a constant term which is also the lower bound for the two right-hand sides in the inequalities (5.7) and (5.8), respectively. It follows that in order to get close and closer to $\ln(\det\{\Sigma_u\})$ the right-hand sides of the last two inequalities must be minimized with respect to the first stage autoregression lag-order and the second stage Kronecker indices parameters, respectively. Furthermore, the right hand-side of the inequalities (5.7) and (5.8) depend on the norm of the estimated eigenvalue matrices as well as the norm of the difference between the

true and the estimated ones. These eigenvalues are the zeros of polynomial functions in their corresponding innovation covariance matrices. Of course, such functions are continuous, so they preserve the rate of convergence stated above between the true and the estimated innovation covariances. Therefore, using the convergence rate results (3.18) and (3.42) stated in Propositions **3.3** and **3.6**, respectively, and the fact that the norm of any positive definite symmetric covariance matrix is equal to that corresponding to its eigenvalues matrix, we have

$$\left\| \Lambda_{\Sigma_u} - \Lambda_{\tilde{\Sigma}_u(n_T)} \right\| = O_p\left(k^2 n_T / T^{1/2}\right), \quad \left\| \Lambda_{\Sigma_u} - \Lambda_{\tilde{\Sigma}_e(n_T)} \right\| = O_p\left(r_{\bar{p}} n_T / T^{1/2}\right). \quad (5.9)$$

The next proposition establishes the consistency of the first-stage selected lag-autoregression order as well as the corresponding second-stage estimated Kronecker indices.

Proposition 5.1 FIRST AND SECOND STAGES ESTIMATED ORDERS CONSISTENCY. *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.8)-(2.11). *Let also n_{T_0} and $P_0 = (p_{1,0}, \ldots, p_{k,0})'$ be the true first-stage lag-autoregression and second stage Kronecker indices that minimize the information criteria*

$$Cr\left(n_T\right) = \ln\left(\det\left\{ \tilde{\Sigma}_u\left(n_T\right) \right\} \right) + c_1 k^2 \frac{n_T}{T^{1/2}} \left(1 + c_1 k^2 \frac{n_T}{T^{1/2}} \right), \quad (5.10)$$

$$Cr\left(P\right) = \ln\left(\det\left\{ \tilde{\Sigma}_e\left(n_T\right) \right\} \right) + c_2 r_{\bar{p}} \frac{n_T}{T^{1/2}} \left(1 + c_2 r_{\bar{p}} \frac{n_T}{T^{1/2}} \right), \quad (5.11)$$

respectively, where c_1 and c_2 are positive constants sufficiently small. Then, under the assumptions **2.1** *to* **2.4**, *the corresponding estimates $\hat{n}_T$ and $\hat{P} = (\hat{p}_1, \ldots, \hat{p}_k)'$ obtained by minimizing such criteria are strongly consistent.*

These information criteria when combined with the shortcuts that we provide herein yield strongly consistent estimates of the dynamic indices. In particular, to simplify the minimization of such criteria in big VARMA systems, we propose shortcuts that provide strongly consistent estimates of the Kronecker indices. These shortcuts

can also be implemented easily using the estimation procedure considered in Dufour and Jouini (2005). Moreover, our proposed minimizing algorithm does not depend on any prespecified upper bound for these dynamic indices. Which seems quite plausible, more especially, as such information could not be available in practice.

The idea is very simple, it starts from a WN process where for a given component of the vector of Kronecker indices we move j steps ahead (by augmenting only the value of the given component by j units, $1 \leq j \leq k$, leaving unchanged the other components), and the criterion $Cr(P)$ is evaluated. Then the procedure is repeated to all k components (directions), separately, with the same value j. Which constitutes here a cycle of length k and magnitude j. Thereafter, at the comparison level, we retain the index for which the reduced criteria over the k directions (if they exist) are minimized. The algorithm is initialized with $j = k$, and the procedure is sequentially repeated starting from the vector of Kronecker indices already retained in the previous step until no further reduction in the criterion is possible for all components. In the latter case, the procedure will be sequentially reinitialized with $j = j - 1$ and so on (for each cycle) as long as no further reductions in the criterion are possible until $j = 1$, in which case the algorithm stops. If however, the criterion is minimized for some j $(1 \leq j \leq k-1)$ the procedure is sequentially reinitialized with $j = k$.

More formally, let

$$P^{(0)} = \left(p_1^{(0)}, \ldots, p_k^{(0)}\right)' \tag{5.12}$$

be the initial vector of Kronecker indices associated with the WN process, that is $p_l^{(0)} = 0$ for $l = 1, \ldots, k$, where (0) signifies that zero cycle has been completed. Then for $i \geq 1$ and $1 \leq l, j \leq k$, define

$$P^{(i-1)} = \left(p_1^{(i-1)}, \ldots, p_{l-1}^{(i-1)}, p_l^{(i-1)}, p_{l+1}^{(i-1)}, \ldots, p_k^{(i-1)}\right)', \tag{5.13}$$

$$P^{(i)}(l,j) = \left(p_1^{(i-1)}, \ldots, p_{l-1}^{(i-1)}, p_l^{(i-1)} + j, p_{l+1}^{(i-1)}, \ldots, p_k^{(i-1)}\right)', \tag{5.14}$$

and let

$$\mathcal{C}^{(i)}(j) = \left\{l : 1 \leq l \leq k ; Cr\left(P^{(i)}(l,j)\right) < Cr\left(P^{(i-1)}\right)\right\}. \tag{5.15}$$

Now, take $i = 1$ and $j = k$. If $\mathcal{C}^{(i)}(j)$ is not empty then the estimated vector of Kronecker indices at the i-th cycle is such that

$$\hat{P}^{(i)} = P^{(i)}(s,j) \Big| Cr\Big(P^{(i)}(s,j)\Big) = \min_{l \in \mathcal{C}^{(i)}(j)} \Big\{ Cr\Big(P^{(i)}(l,j)\Big) \Big\} \tag{5.16}$$

and the procedure is sequentially repeated with $i = i + 1$, $P^{(i-1)} = \hat{P}^{(i)}$ and $j = k$. Contrarily, if $\mathcal{C}^{(i)}(j)$ is empty then $\hat{P}^{(i)} = P^{(i-1)}$, $i = i + 1$, $P^{(i-1)} = \hat{P}^{(i)}$ and $j = j - 1$. In the latter case, if the set $\mathcal{C}^{(i)}(j)$ becomes not empty however for some j $(1 \leq j \leq k - 1)$ then $i = i + 1$, $P^{(i-1)} = \hat{P}^{(i)}$ and $j = k$, and if instead, it remains empty until $j = 1$ then the algorithm stops with $\hat{P} = \hat{P}^{(i)}$. The key feature of this procedure is that it is able to skip a local minimum whenever another lowest minimum is available at least in the nearest promising area. Hence, our procedure ought to find the local minimum which is most likely to be the global one, from the parsimony viewpoint of VARMA models. Moreover, as T grows to infinity, higher values of j, say $1 < j \leq k - 1$, are not needed to ensure the convergence of the estimated Kronecker indices to their true values.

So far, the estimated echelon form VARMA models (in the second and the third stages) are assumed to be stationary and invertible. In practice, this could not be the case since we perform linear regressions without imposing such constraints. Which yields third-stage regressors that may explode namely in moderate and large sample sizes [see (4.22)]. To obtain an invertible moving average operator Hannan and Kavalieris (1984b) suggested new echelon form moving average coefficients such as

$$\tilde{\Theta}^{\lambda}(z) = \tilde{\Theta}(0) + \lambda\Big[\tilde{\Theta}(z) - \tilde{\Theta}(0)\Big], \; 0 < \lambda \leqq 1 \tag{5.17}$$

where $\det\{\tilde{\Theta}(z)\} = 0$ for some z with $|z| < 1$. They have proposed choosing λ as near to unity as $\tilde{\Theta}^{\lambda}(z)$ is stable, and in case the zeros of $\tilde{\Theta}(z)$ were very close to the unit circle (*i.e.* smaller than 1.01 in modulus) they have recommended selecting λ so that the zeros of $\tilde{\Theta}(z)$ will be far from the unit circle. Hence, causing the forcing initialization effects to die a little more rapidly. Note that such technique rests on

values of λ that could be chosen arbitrarily. Therefore, in view of (5.17) it seems noisy for applied researchers to choose the appropriate value of λ that ensures stationarity and invertibility constraints as well since

$$\tilde{\Phi}_0^{\lambda} = \tilde{\Phi}_0 + (1-\lambda)\sum_{j=1}^{\bar{p}} \tilde{\Theta}_j \tag{5.18}$$

might induce a nonstationary echelon form model even if it was not, previously.

In the following we provide an efficient algorithm that completely overcomes such situations, as it picks the appropriate value of λ given the considerations taken by the practitioners (focusing on stationarity or/and invertibility conditions). The algorithm we describe here for the second stage estimation could also be considered to obtain stationary invertible third-stage echelon form estimates. Moreover, it can be easily adapted in the first stage of the estimation procedure to estimate a stationary long VAR.

To this end, define

$$\tilde{A}(L) = I_k - \sum_{i=1}^{\bar{p}} \tilde{A}_i L^i, \quad \tilde{B}(L) = I_k + \sum_{j=1}^{\bar{p}} \tilde{B}_j L^j \tag{5.19}$$

where $\tilde{A}_i = \tilde{\Phi}_0^{-1}\tilde{\Phi}_i$ and $\tilde{B}_j = \tilde{\Phi}_0^{-1}\tilde{\Theta}_j$ for $i, j = 1, \ldots, \bar{p}$, with $\tilde{\Phi}_i$ and $\tilde{\Theta}_j$ the second-stage estimates of Φ_i and Θ_j, respectively. Then compute the eigenvalues of the following two $(k\bar{p}) \times (k\bar{p})$ matrices

$$\tilde{A} = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 & \cdots & \tilde{A}_{\bar{p}-1} & \tilde{A}_{\bar{p}} \\ I_k & O_k & \cdots & O_k & O_k \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \cdots & \ddots & O_k & O_k \\ O_k & \cdots & \cdots & I_k & O_k \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} -\tilde{B}_1 & -\tilde{B}_2 & \cdots & -\tilde{B}_{\bar{p}-1} & -\tilde{B}_{\bar{p}} \\ I_k & O_k & \cdots & O_k & O_k \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \cdots & \ddots & O_k & O_k \\ O_k & \cdots & \cdots & I_k & O_k \end{bmatrix} \tag{5.20}$$

to check whether the estimated model of (3.19) is stationary and/or invertible or not. Now, let $\lambda_i(\tilde{A})$ and $\lambda_j(\tilde{B})$ be the corresponding eigenvalues associated with $\tilde{A}$ and $\tilde{B}$,

respectively. Set also

$$\lambda_{\tilde{A}} = \max\left\{|\lambda_1(\tilde{A})|, \ldots, |\lambda_{\bar{p}}(\tilde{A})|\right\}, \quad \lambda_{\tilde{B}} = \max\left\{|\lambda_1(\tilde{B})|, \ldots, |\lambda_{\bar{p}}(\tilde{B})|\right\}, \tag{5.21}$$

$$\lambda_{\max} = \max\left\{\lambda_{\tilde{A}}, \lambda_{\tilde{B}}\right\}, \quad \lambda_{\min} = \min\left\{\lambda_{\tilde{A}}, \lambda_{\tilde{B}}\right\}. \tag{5.22}$$

Then, if $\lambda_{\max} \geq 1$, define

$$\lambda_{star} = \begin{cases} \lambda_{\max} & \text{if } \lambda_{\min} < 1, \\ \lambda_{\min} & \text{if } \lambda_{\min} \geq 1. \end{cases} \tag{5.23}$$

Further, consider the iterated algorithm such as

$$\tilde{\Phi}_0^{(l)} = \left(c\lambda_{star}^{(l-1)}\right)^{-1} \tilde{\Phi}_0^{(l-1)} \quad \text{if } \lambda_{star} \geq 1\,, \tag{5.24}$$

$$\tilde{A}_i^{(l)} = \left(c\lambda_{\tilde{A}}^{(l-1)}\right)^{-1} \tilde{A}_i^{(l-1)} \quad \text{if } \lambda_{\tilde{A}} \geq 1\,, \quad i = 1, \ldots, \bar{p}, \tag{5.25}$$

$$\tilde{B}_j^{(l)} = \left(c\lambda_{\tilde{B}}^{(l-1)}\right)^{-1} \tilde{B}_j^{(l-1)} \quad \text{if } \lambda_{\tilde{B}} \geq 1\,, \quad j = 1, \ldots, \bar{p} \tag{5.26}$$

where (l) designates the l-th iteration ($l \geq 1$), c is a positive constant greater than one to avoid models with unit roots, $\tilde{\Phi}_0^{(0)} = \tilde{\Phi}_0$, $\tilde{A}_i^{(0)} = \tilde{A}_i$, $\tilde{B}_j^{(0)} = \tilde{B}_j$, $\lambda_{star}^{(0)} = \lambda_{star}$, $\lambda_{\tilde{A}}^{(0)} = \lambda_{\tilde{A}}$ and $\lambda_{\tilde{B}}^{(0)} = \lambda_{\tilde{B}}$. Note that reasonable values of c equal 1.01 or 1.02 can be used to ensure the convergence of the proposed algorithm. Specifically, $\lambda_{\tilde{A}}^{(l-1)}$, $\lambda_{\tilde{B}}^{(l-1)}$ and $\lambda_{star}^{(l-1)}$ are such that

$$\lambda_{\tilde{A}}^{(l-1)} = \max\left\{|\lambda_1(\tilde{A}^{(l-1)})|, \ldots, |\lambda_{\bar{p}}(\tilde{A}^{(l-1)})|\right\}, \tag{5.27}$$

$$\lambda_{\tilde{B}}^{(l-1)} = \max\left\{|\lambda_1(\tilde{B}^{(l-1)})|, \ldots, |\lambda_{\bar{p}}(\tilde{B}^{(l-1)})|\right\}, \tag{5.28}$$

where $\tilde{A}^{(l-1)}$ and $\tilde{B}^{(l-1)}$ are formed as in (5.20) with the $\tilde{A}_i^{(l-1)}$'s and the $\tilde{B}_j^{(l-1)}$'s are given in (5.25) and (5.26). If, furthermore, $\lambda_{\max}^{(l-1)} \geq 1$, then we define again

$$\lambda_{star}^{(l-1)} = \begin{cases} \lambda_{\max}^{(l-1)} & \text{if } \lambda_{\min}^{(l-1)} < 1, \\ \lambda_{\min}^{(l-1)} & \text{if } \lambda_{\min}^{(l-1)} \geq 1, \end{cases} \tag{5.29}$$

with

$$\lambda_{\max}^{(l-1)} = \max\left\{\lambda_{\tilde{A}}^{(l-1)}, \lambda_{\tilde{B}}^{(l-1)}\right\}, \quad \lambda_{\min}^{(l-1)} = \min\left\{\lambda_{\tilde{A}}^{(l-1)}, \lambda_{\tilde{B}}^{(l-1)}\right\}, \tag{5.30}$$

and we proceed likewise until $\lambda_{\tilde{A}}^{(l)}$ and $\lambda_{\tilde{B}}^{(l)}$ become both less than one, then the algorithm stops and we define $\tilde{\Phi}_0^* = \tilde{\Phi}_0^{(l)}$, $\tilde{A}_i^* = \tilde{A}_i^{(l)}$, $\tilde{B}_j^* = \tilde{B}_j^{(l)}$, $\tilde{\Phi}_i^* = \tilde{\Phi}_0^* \tilde{A}_i^*$ and $\tilde{\Theta}_j^* = \tilde{\Phi}_0^* \tilde{B}_j^*$ for $i, j = 1, \ldots, \bar{p}$. In particular, to obtain ML estimates, the maximizing likelihood algorithm can be initialized with

$$\tilde{\eta}^* = R' \mathrm{vec}\left[\tilde{\mu}_{\Phi}^*, I_k - \tilde{\Phi}_0^*, \tilde{\Phi}_1^*, \ldots, \tilde{\Phi}_{\bar{p}}^*, \tilde{\Theta}_1^*, \ldots, \tilde{\Theta}_{\bar{p}}^*\right] \tag{5.31}$$

where $\tilde{\mu}_{\Phi}^* = \left(\tilde{\Phi}_0^* - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i^*\right)\tilde{\mu}_y$. It is worth noting that this algorithm, satisfying the stationarity and invertibility constraints as well, can be easily and efficiently used to develop an improved maximizing likelihood algorithm. Which looks interestingly appealing for obtaining ML estimates in stationary invertible VARMA models whatever is the identification issue.

6. Simulation study

To evaluate the performance of the proposed estimation procedure we consider a simulation study with different sample sizes (T=100, 200, 300, 400, 500, 1000, 2000). Several examples of identified VARMA models under the echelon form representation for different dimensions and Kronecker indices values, are also considered. The simulated models are generated using (2.8) with Gaussian errors, initiating with $y_{t-i}, u_{t-i} = 0$ for all $t \leq i$, $i = 1, ...\bar{p}$, and such that $u_t = P_u \varepsilon_t$ where P_u is a lower triangular matrix satisfying $P_u P_u' = \Sigma_u$ and $\varepsilon_t \sim N[0, I_k]$.[1] Simulation results for the frequencies of the estimated Kronecker indices as well as the finite sample characteristics of the echelon form VARMA parameter estimates, such as mean, average-deviation (Avg-Dev), median, 5^{th}-percentile, 95^{th}-percentile, root mean square errors (RMSE) and standard-deviation (Std-Dev), are given in tables xi through xxvi. Theses results are obtained over 1000 trials for each model using GAUSS random number generator

[1] The values of such a matrix for all simulated models are defined in Appendix B.

(version 3.2.37). To avoid numerical problems in the MC simulation study, that may be caused by the initializing effect, we dropped the first 100 simulated pseudo data.

In our procedure, for all models in all replications the long autoregression lag order is selected by minimizing the information criterion (5.10) with $c_1 = 0.10\sqrt{2}/\sqrt{k}$, whereas the Kronecker indices are estimated using the proposed shortcuts as described in (5.13) - (5.15) by minimizing the information criterion (5.11) with $c_2 = 0.10\sqrt{2}/k$, given the selected autoregression-lag in the first-stage. What is relevant to stress here is that the penalty adjustment in each stage (the first two steps) of the estimation procedure must be moderately smaller for highly dimensional systems to avoid excessive less parameterized models when it should not be the case. At each stage along with the estimation procedure, stationarity (for the first, second and third stages) and invertibility (for the second and third stages) constraints are imposed through (5.19) - (5.28). However in the cases where the satisfaction of these constraints leads to models with likelihood too far from the maximum we proposed as consistency criteria

$$\ln\left(\det\left\{\tilde{\Sigma}_u\left(\hat{\eta}\right)\right\}\right) < \ln\left(\det\left\{\tilde{\Sigma}_u\left(n_T\right)\right\}\right) + \frac{c_0}{\sqrt{T}} \tag{6.1}$$

and

$$\ln\left(\det\left\{\tilde{\Sigma}_u\left(\hat{\eta}\right)\right\}\right) < \ln\left(\det\left\{\tilde{\Sigma}_u\left(\tilde{\eta}\right)\right\}\right) \tag{6.2}$$

when interested in the third step estimates, or

$$\ln\left(\det\left\{\tilde{\Sigma}_u\left(\tilde{\eta}\right)\right\}\right) < \ln\left(\det\left\{\tilde{\Sigma}_u\left(n_T\right)\right\}\right) + \frac{c_0}{\sqrt{T}} \tag{6.3}$$

and

$$\ln\left(\det\left\{\tilde{\Sigma}_u\left(\tilde{\eta}\right)\right\}\right) < \ln\left(\det\left\{\tilde{\Sigma}_e\left(n_T\right)\right\}\right) \tag{6.4}$$

when only interested in the second step estimates, where c_0 is a positive constant sufficiently small.[2] The criteria (6.1) and (6.2), and (6.3) and (6.4) ensure that we are maximizing the VARMA likelihood as far as possible by performing only three linear

[2]To have an idea how to choose the value of such a constant you have just to fix its value in accordance with the quantity $\ln(\det\{\tilde{\Sigma}_u(n_T)\})$.

regressions or two linear regressions, respectively, without using the ML estimation procedure. These criteria can also be combined with the above algorithm by considering in the first stage of the estimation procedure, increasing then decreasing lags up to five units until all the required constraints could be met. Therefore, the whole algorithm stops. The number of five lags is arbitrarily. The idea behind this is that the true lag order associated with the long autoregression can be easily met between these two extreme values of lag orders as the sample size grows. If for some trials the algorithm fails to provide an estimated model satisfying all the above requirements, the corresponding simulated pseudo data are dropped then replaced with other ones. This is what we have designated in tables xviii, xxi, xxii, xxv and xxvi by M2 as a reference to the second method. Although the frequency of replaced trials for each model is small (around 1% to 15% out of 1000 trials with T=100) and decreases with the sample size even in highly persistent models, we proposed alternative algorithm to minimize it. Such an algorithm is similar to M2 but instead of replacement of the pseudo data, in case the algorithm fails to provide stationary invertible VARMA model with likelihood as high as possible, we rather consider M2 with decreasing values of c_2 beginning from the above chosen value (see page 119) to a sufficiently small one. Such value can be chosen so that to avoid excessive overparameterized models that may easily induce nonstationarity and/or noninvertibility, hence nonregularity problems. Again, if for some trials and for different values of c_2 the algorithm fails as in M2 then the pseudo data is replaced. Of course in this case the number of replaced trials for each model will be smaller. This is what we have also designated in tables xviii, xxi, xxii, xxv and xxvi by M1 as a reference to the first method. Note that the simulation results presented in tables xi - xvii are obtained using M1.

To give an idea how accurate is our method in providing high frequencies for the estimated true Kronecker indices characterizing the echelon form VARMA models, we have also considered a simulation study in which we compared our results to those obtained by Hannan and Kavalieris (1984b) and Poskitt (1992) methods. These results are presented in tables xviii, xxi, xxii, xxv and xxvi. A good description of the shortcuts as well as the information criteria they proposed in their respective

methods is given in Lütkepohl (1991, Chapter 8, Subsections 8.3.3 and 8.3.4). The algorithm used in these methods for estimating the lag-order of the long autoregression and the Kronecker indices as well, is basically the same as for M2 of our method with the main difference that such an algorithm is based on information criteria and shortcuts suggested respectively within each approach. The selected lag-order for the long autoregression in each approach is obtained by minimizing the information criteria

$$\ln\big(\det\{\tilde{\Sigma}_u(n_T)\}\big) + k^2 n_T \frac{\ln T}{T} \tag{6.5}$$

and

$$\ln\big(\det\{\tilde{\Sigma}_u(n_T)\}\big) + k^2 n_T \frac{2}{T} \tag{6.6}$$

for M1 and M2, respectively. To estimate the Kronecker indices characterizing the echelon form VARMA models with Hannan and Kavalieris (1984b) approach, these indices are however obtained using shortcuts for the second and third stages by minimizing

$$\ln\big(\det\{\tilde{\Sigma}_e(n_T)\}\big) + r_{\bar{p}} \frac{2\ln\ln T \ln T}{T} \tag{6.7}$$

for M1 and

$$\ln\big(\det\{\tilde{\Sigma}_e(n_T)\}\big) + r_{\bar{p}} \frac{\ln T}{T} \tag{6.8}$$

for M2. For Poskitt (1992) approach such indices are estimated using shortcuts based on minimized single equation information criteria. These individual information criteria are

$$\ln\big(\det\{\tilde{\sigma}_{e,i}(n_T)\}\big) + r_{\bar{p},i} \frac{2\ln\ln T \ln T}{T}, \quad i = 1, \ldots k \tag{6.9}$$

for M1 and

$$\ln\big(\det\{\tilde{\sigma}_{e,i}(n_T)\}\big) + r_{\bar{p},i} \frac{\ln T}{T}, \quad i = 1, \ldots k \tag{6.10}$$

for M2, where $\tilde{\sigma}_{e,i}(n_T)$ is the estimated standard error associated with the i-th equation and $r_{\bar{p},i}$ the number of freely varying parameters implied by any vector of Kronecker indices $P = (p_1, \ldots, p_k)'$ on the i-th equation.

From tables xi and xii, simulation results show that our method performs very

well, as it provides high frequencies of estimating the true Kronecker indices. In almost all models these frequencies grow fast to reach at least 90% in moderate sample sizes. Moreover, our approach looks stable in the sense that such estimated probabilities are nondecreasing functions of the sample size. Also, based on the estimated frequencies of the Kronecker indices given in the above tables, simulation results with some picked examples of bivariate ARMA models on the finite sample characteristics of the echelon form parameter estimates, show good properties in small and moderate sample sizes. As we can see from tables xiii - xvii, such estimates converge quickly to their true values. Their average deviations from their corresponding true values are small. In all models, for almost all coefficients these deviations are around two and three digits for the second and third stages estimates, respectively. In addition, based on the median, the 5^{th} and 95^{th} percentiles, a prior look at the sample distribution of such estimates over 1000 trials could be thought as normal. The RMSE associated with each parameter is relatively small and in all cases is greater or equal to the Std-Dev.[3] Moreover, simulation results show always the gain in efficiency in performing the third-step estimation since there is no substantial cost in running an additional linear regression. Indeed, as shown in tables xiii - xvii, the finite sample characteristics of the third-stage estimates are better that those associated with the second-stage estimates since we can report for example Std-Dev smaller about 44%.[4]

Although our proposed method seems very satisfactory for several bivariate ARMA models, in view of the simulation results drawn in tables xi through xvii, we have also considered VARMA models with higher dimensions. The main finding that should be strongly highlighted in bold from a comparative simulation study, gives supports and provides more credibility to our method in estimating the true Kronecker indices against Hannan and Kavalieris (1984b) and Poskitt (1992) methods as well, especially, in higher dimensional VARMA systems. In fact, tables xviii, xxi, xxii, xxv and xxvi, show higher estimated frequencies of the true Kronecker indices using both variants of our methods for all picked models. Such estimated frequencies are

[3]Based on the first fourth digits we reported here in the tables.

[4]See the bottom corners of the two panels of table xv.

non decreasing functions with the sample size. Moreover, they imply probabilities as high as 95% with increasing sample sizes. However, the results associated with Hannan and Kavalieris (1984b) and Poskitt (1992) methods are very poor for both of their variants in all models. They fail to correctly estimate such indices with substantial probabilities. Moreover, in view of our simulation study these methods can be unreliable since they provide estimated frequencies of the Kronecker indices that are non stable with the sample size and can behave in a brutal way. As an example, with M1 of Hannan and Kavalieris (1984b) method we can see from table xxii a probability of estimating the true Kronecker indices $P = (3, 1, 2)$ of about 0.4% with T=1000 then a probability of about 92.7% for T=2000. The same thing happens with M1 of Poskitt (1992) method as table xxv shows a probability of about 13.7% to 73.8% for T=1000 to T=2000 in estimating the true Kronecker indices $P = (3, 2, 4, 1)$. From another simulated example table xxvi shows also that alternative methods can fail dramatically to estimate the true Kronecker indices, while our proposed method provides very good results. So one cannot be sure when implementing these procedures to estimate consistently the true VARMA model with a substantial probability within the echelon form identification framework. Finally in tables xix, xx, xxiii and xxiv we only report the finite sample characteristics of the echelon form VARMA parameter estimates obtained using our method, as those provided respectively by Hannan and Kavalieris (1984b) and Poskitt (1992) procedures are very poor. So to save space we dropped them. Again, these finite sample properties look nice given the considered sample sizes and the increasing number of the echelon form VARMA coefficients implied in these two examples.

7. Empirical illustration

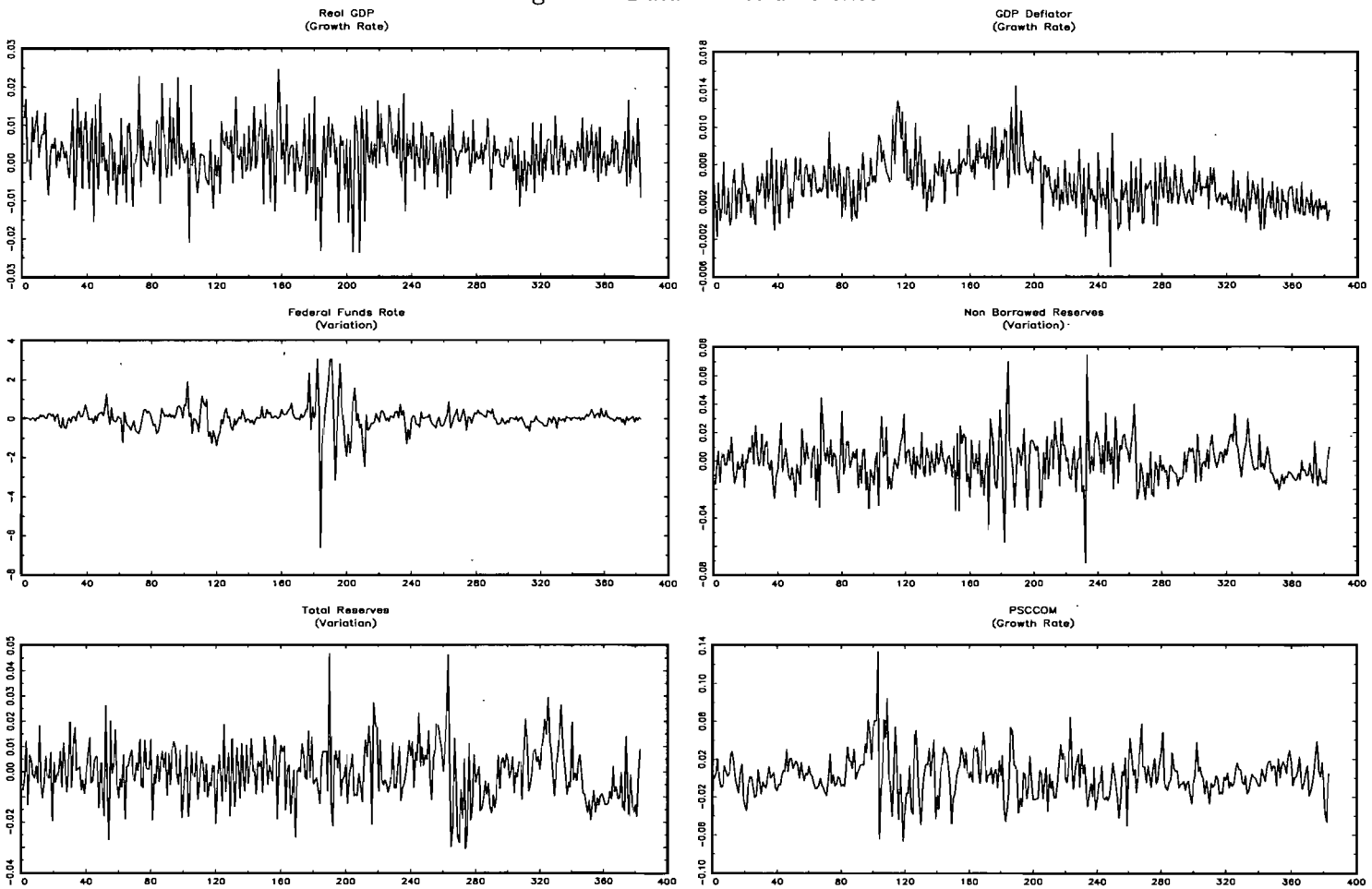
To illustrate our estimation method we consider measuring and identifying U.S. monetary policy within the echelon VARMA model framework. To show the advantage of using such parsimonious models compared to VAR models we study the implied impulse response functions (IRFs) from a system of six macroeconomic time series.

Recently, VAR models have been widely used to analyze and estimate the effect of monetary policy shocks [see, for example, Bernanke and Blinder (1992), Christiano and Eichenbaum (1992), Gordon and Leeper (1994), Lastrapes and Selgin (1995), Strongin (1995), Pagan and Robertson (1995, 1998), Christiano, Eichenbaum, and Evans (1994, 1996, 1999), Bernanke and Mihov (1998), Koray and McMillin (1999), McMillin (2001), Kim and McMillin (2003) and Lastrapes and McMillin (2004)]. Two methods have been widely used in the VAR literature to identify structural shocks to monetary policy. The first one puts restrictions on the contemporaneous relations among the variables of the VAR system [see, for example, Bernanke and Blinder (1992), Strongin (1995), Christiano, Eichenbaum, and Evans (1994, 1996, 1999), Bernanke and Mihov (1998)], while the second one imposes restrictions on the long-run relations among the variables [Koray and McMillin (1999), McMillin (2001), Kim and McMillin (2003), Lastrapes and McMillin (2004)]. Although such restrictions can be used to rationalize each approach using economic as well institutional arguments, there is no consensus about which one is preferable. In what follows, we consider the second identification scheme by assuming the long-run neutrality of monetary policy as in McMillin (2001) without imposing contemporaneous restriction among the variables, then evaluating the effects of monetary policy shocks by computing the IRFs.

The data used here come from Bernanke and Mihov (1998). These data are monthly and cover the period January 1962 to December 1996. The use of monthly data reduces problems that may arise with temporal aggregation in estimating models [see Christiano and Eichenbaum (1987)]. They consist of total bank reserves, nonborrowed reserves and the federal funds rate. The non-policy variables - real GDP, GDP deflator and the Dow-Jones index of spot commodity (PSCCOM) were taken in log. Further, the policy variables total bank reserves and nonborrowed reserves were normalized as in Bernanke and Mihov (1998) by a long (36-month) moving average of total bank reserves to induce stationarity. As shown in Figure 1, the plotted series in first difference reveals stationary pattern.

Following the same arguments as in McMillin (2001), we fit a VAR(12) with a

Figure 1 : Data in first difference



constant to the ordered series in first difference : log of real GDP, the log of real commodity prices (PSCCOM/GDP deflator), real funds rate, total bank reserves and nonborrowed reserves. Such an ordering of the variables is suggested by the Cholesky decomposition of the long-run relations of the innovations covariance matrix to easily impose neutrality and hence identify the structural shocks to monetary policy measured here by nonborrowed reserves [see Keating (2002)]. Using a VAR(12) in the first stage of our estimation procedure we estimate an echelon form VARMA model with Kronecker indices $P = (5, 4, 2, 2, 3, 4)$ and a constant term. A description of the estimated model is provided in Appendix C. Using the portmanteau test statistic, the null hypothesis of white noise process of the estimated residuals associated with this model cannot be rejected at nominal level of 5% as we can report a nonparametric bootstrap p-value of 22.17%. The following panel provides the contemporaneous effects of one standard deviation of the structural shocks to nonborrowed reserves (measured as monetary policy variable) on the macroeconomic variables, as implied by the estimated models; VAR(12) in McMillin (2001), our estimated VAR(12) as well as our estimated echelon form VARMA, respectively.

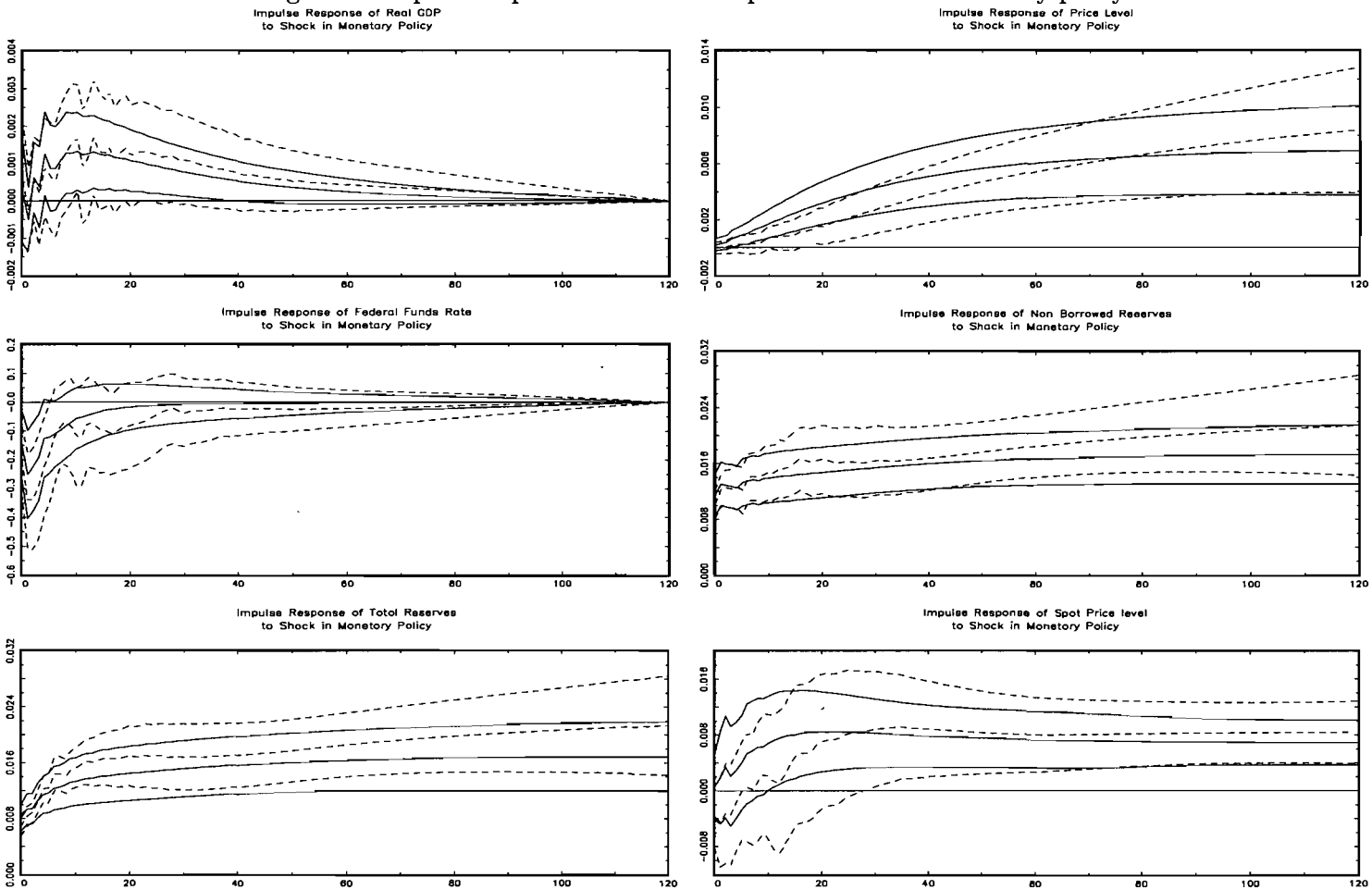
Long-Run Monetary Neutrality Restriction	McMillin	Our	Method
Variable	VAR	VAR	VARMA
Real GDP	0.00016	0.00048	0.00018
GDP deflator	-0.00002	-0.00000	0.00020
PSCCOM	-0.0023	-0.0036	0.00057
Nonborrowed Reserves	0.0101	0.0102	0.0115
Total Bank Reserves	0.0072	0.0070	0.0081
Federal Funds Rate	-0.2028	-0.1906	-0.1470

As we can see from the above results the contemporaneous impact effects of monetary policy shocks on the macroeconomic variables are quite similar for both of the estimated VAR models For example the impact effects of one standard deviation to shocks in nonborrowed reserves on federal funds rate are about -0.2028 and -0.1906 in

McMillin's VAR(12) and our VAR(12), respectively. It is worthnoting that the impact effect on such variable is less in magnitude (-0.1470) when we consider the estimated echelon VARMA model. Further, the sign of the contemporaneous effect of nonborrowed reserves on GDP deflator and PSCCOM is positive from the estimated VARMA model, while it is negative for both estimated VARs (puzzle effect). Given this, one could expect that the timing and the magnitude of the effects of the monetary policy actions would be sensibly different when we consider a VAR(12) compared to the more parsimonious echelon VARMA model with Kronecker indices $P = (5, 4, 2, 2, 3, 4)$. For that we compute the implied IRFs for each model. Since the fitted models are in first difference the implied moving average representations indicate the effect of shocks on the changes in the variables. Hence the IRFs at a given horizon are the cumulative shocks up to that horizon. These IRFs are plotted in figure 2.

This figure rises some differences in the behavior of the IRFs generated by the VAR model compared to those obtained from the VARMA model. In fact, although the IRFs of output (the log of real GDP) to structural shocks to nonborrowed reserves are closely similar for both models at the shortest horizons, those computed from the VARMA model, instead, decline earlier and more rapidly (after reaching a peak) toward the initial level, with increasing horizons. For the price level (log of the GDP deflator) these IRFs stabilize more quickly in the VARMA case to reach a higher level than initially but lower to what would be in the VAR case. Moreover, for the federal funds rate we report a smaller initial impact in the first month of about -0.25 from the VARMA model versus -0.32 from the VAR case. Again, after an initial decrease the federal funds rate goes up to reach its initial level more quickly in the VARMA model. Although these results show the long-run neutrality of monetary policy in both models, we also show that the IRFs of all variables to structural shocks to monetary policy stabilize more quickly in the VARMA model compared to the VAR case. Interestingly, we can also see from Figure 2 that the estimated echelon VARMA model tends to deliver smoother IRFs compared to those generated by the estimated long VAR. Moreover, the confidence band corresponding to the IRFs are more tighter in the VARMA case with amplitudes that tend to stabilize fast than in

Figure 2 : Impulse response functions to a positive shock in monetary policy



The dashed lines correspond to VAR(12) model while the solid lines correspond to echelon VARMA model with Kronecker indices (5,4,2,2,3,4). The confidence bands are about one standard deviation. The standard deviations are computed using Monte Carlo integration with 1000 draws.

VAR model. Hence, in view of these results, the estimated VARMA model in echelon form looks more precise in measuring the magnitude and the spread of the monetary policy effects. This is not surprising as the VARMA models are more parsimonious than VAR models, so they tend to deliver more precise parameters estimates and consequently more precise IRFs.

8. Conclusion

In this paper, we have proposed a three-step linear estimation procedure for estimating stationary invertible echelon form VARMA models. The method can easily be adapted to VARMAX models and extended to integrated and cointegrated VARMA models as well. The estimation method focuses on the echelon form as it tends to deliver relatively parsimonious parameterized models. But our procedure remains valid to other identifying issues such as final equations or any restricted model for inference purposes.

Our proposed method provides a simplified general standard form for the echelon form parameter estimates that are more easy to compute than those of Hannan and Kavalieris (1984b) and Poskitt (1992) respectively. Which may look more advisable and tractable from the practical viewpoint. Further, we have extended the results of Dufour and Jouini (2005) for the two-step estimation method to derive the asymptotic distribution of the GLS estimators in case of strong white noises, since it has not been stated yet anywhere. We have also provided the theoretical justification for implementing a third stage linear regression to allow for efficiency in case of Gaussian errors. This shows the asymptotic equivalence between the distribution of our third-stage estimates and that of ML estimators. The finite sample properties of our estimates are not the same as those of ML estimators, however. Although our third-stage estimation procedure is equivalent to that of Hannan and Kavalieris (1984b), the estimates of the asymptotic covariances of the echelon form parameters we have given for the second and third stages as well, are simple, nice and easy to use for inference purposes.

In this work, we have also proposed a simplified order selection procedure for identifying the Kronecker indices characterizing the echelon form VARMA representation. Our procedure rests on determining the implied matrix of restrictions for all possible sets of Kronecker indices for any given dimension of VARMA systems. Thus, by such an achievement we have made the echelon form more attractive for further applied works. Further, to solve the overparameterization and accuracy problems missed, respectively, in Hannan and Kavalieris (1984b) and Poskitt (1992) methods, we have proposed information criteria and shortcuts to provide strongly consistent estimates for the true Kronecker indices in moderate and large sample framework. Furthermore, to ensure stationary invertible estimated echelon form VARMA models using our estimation procedure, we have developed an efficient algorithm that works in a systematic way to that end. Such an algorithm could be easily integrated to improve the ML maximization packages for estimating stationary invertible VARMA models. Moreover, a simulation study shows that our estimation method dominates so far those of Hannan and Kavalieris (1984b) and Poskitt (1992), respectively. This is so, especially and even more in higher dimensional systems. That is, unlike to their approaches which behave poorly for estimating the Kronecker indices, hence providing bad finite sample properties for the implied echelon form parameter estimates, our procedure, instead, yields high probabilities for identifying the true echelon form VARMA model. Moreover, it tends to deliver nice and enough satisfactory finite sample properties for the corresponding parameter estimates. So, in view of the above results our proposed estimation method looks more accurate and more advisable in practice.

Finally, through an empirical example on the U.S. economy with six macroeconomic series we have easily estimated an echelon form VARMA model using our proposed method. Further, we have shown that these more parsimonious models provide better and efficient tools compared to the VARs in analyzing some economic policies as in the case of monetary policy studied herein.

A. Appendix : Proofs

PROOF OF PROPOSITION **3.1** Note first that

$$\mathsf{E}\left\|\tilde{\Gamma}_Y(n_T) - \Gamma_Y(n_T)\right\|_1^2 \leq \mathsf{E}\left\|\tilde{\Gamma}_Y(n_T) - \Gamma_Y(n_T)\right\|^2 \tag{A.1}$$

where $\tilde{\Gamma}_Y(n_T) = T^{-1}\sum_{t=1}^{T} Y_t(n_T)\, Y_t(n_T)'$ and $\Gamma_Y(n_T) = \mathsf{E}\left[Y_t(n_T)\, Y_t(n_T)'\right]$. Then it can be easily seen that

$$\begin{aligned}\mathsf{E}\left\|\tilde{\Gamma}_Y(n_T) - \Gamma_Y(n_T)\right\|^2 &= 2\sum_{\tau=1}^{n_T}\mathsf{E}\left\|T^{-1}\sum_{t=1}^{T}\left(y_{t-\tau} - \mu_y\right)\right\|^2 \\ &\quad + \sum_{\tau_1=1}^{n_T}\sum_{\tau_2=1}^{n_T}\mathsf{E}\left\|T^{-1}\sum_{t=1}^{T}\left[y_{t-\tau_1}y'_{t-\tau_2} - \Gamma_y(\tau_1-\tau_2)\right]\right\|^2,\end{aligned} \tag{A.2}$$

with $\Gamma_y(\tau_1-\tau_2) = E\left[y_{t-\tau_1}y'_{t-\tau_2}\right]$. Using (2.3) it follows from Hannan (1970, Chapter 4) that

$$\mathsf{E}\left\|T^{-1}\sum_{t=1}^{T}\left(y_{t-\tau} - \mu_y\right)\right\|^2 = O_p\left(\frac{k}{T}\right), \tag{A.3}$$

$$\mathsf{E}\left\|T^{-1}\sum_{t=1}^{T}\left[y_{t-\tau_1}y'_{t-\tau_2} - \Gamma_y(\tau_1-\tau_2)\right]\right\|^2 = O_p\left(\frac{k^2}{T}\right), \tag{A.4}$$

hence

$$\mathsf{E}\left\|\tilde{\Gamma}_Y(n_T) - \Gamma_Y(n_T)\right\|^2 = O_p\left(\frac{kn_T}{T}\right) + O_p\left(\frac{k^2n_T^2}{T}\right) = O_p\left(\frac{k^2n_T^2}{T}\right). \tag{A.5}$$

Further, we have

$$\begin{aligned}\left\|\tilde{\Gamma}_Y(n_T)^{-1} - \Gamma_Y(n_T)^{-1}\right\| &= \left\|\tilde{\Gamma}_Y(n_T)^{-1}\left[\tilde{\Gamma}_Y(n_T) - \Gamma_Y(n_T)\right]\Gamma_Y(n_T)^{-1}\right\| \\ &\leq \left\|\tilde{\Gamma}_Y(n_T)^{-1}\right\|\left\|\tilde{\Gamma}_Y(n_T) - \Gamma_Y(n_T)\right\|\left\|\Gamma_Y(n_T)^{-1}\right\|,\end{aligned} \tag{A.6}$$

with

$$\left\|\tilde{\Gamma}_Y(n_T)^{-1}\right\| \leq \left\|\Gamma_Y(n_T)^{-1}\right\| + \left\|\tilde{\Gamma}_Y(n_T)^{-1} - \Gamma_Y(n_T)^{-1}\right\| \tag{A.7}$$

where as in the univariate case $\mathsf{E}\left\|\Gamma_Y(n_T)^{-1}\right\|$ is uniformly bounded by a positive constant for all n_T and

$$\left\|\Gamma_Y(n_T)^{-1}\right\| = O_p(1); \tag{A.8}$$

see Berk (1974, page 491). Moreover, $\left\|\tilde{\Gamma}_Y(n_T) - \Gamma_Y(n_T)\right\|\left\|\Gamma_Y(n_T)^{-1}\right\| < 1$ (an event whose probability converges to one as $T \to \infty$). Therefore

$$\left\|\tilde{\Gamma}_Y(n_T)^{-1} - \Gamma_Y(n_T)^{-1}\right\| \leq \frac{\left\|\tilde{\Gamma}_Y(n_T) - \Gamma_Y(n_T)\right\|\left\|\Gamma_Y(n_T)^{-1}\right\|^2}{1 - \left\|\tilde{\Gamma}_Y(n_T) - \Gamma_Y(n_T)\right\|\left\|\Gamma_Y(n_T)^{-1}\right\|} = O_p\left(\frac{kn_T}{T^{1/2}}\right) \tag{A.9}$$

and finally

$$\left\|\tilde{\Gamma}_Y(n_T)^{-1} - \Gamma_Y(n_T)^{-1}\right\|_1 = O_p\left(\frac{kn_T}{T^{1/2}}\right). \tag{A.10}$$

□

PROOF OF THEOREM **3.1** Recall that $\tilde{\Pi}(n_T) = \tilde{W}_Y(n_T)\tilde{\Gamma}_Y(n_T)^{-1}$, where $\tilde{W}_Y(n_T) = T^{-1}\sum_{t=1}^{T} y_t Y_t(n_T)'$ and $y_t = \Pi(n_T)Y_t(n_T) + u_t(n_T)$. Set

$$U_1(n_T) = T^{-1}\sum_{t=1}^{T}\left[u_t(n_T) - u_t\right]Y_t(n_T)', \quad U_2(n_T) = T^{-1}\sum_{t=1}^{T} u_t Y_t(n_T)'. \tag{A.11}$$

Then

$$\left\|\tilde{\Pi}(n_T) - \Pi(n_T)\right\| \leq \left\{\|U_1(n_T)\| + \|U_2(n_T)\|\right\}\left\|\tilde{\Gamma}_Y(n_T)^{-1}\right\| \tag{A.12}$$

where, by Assumption **2.3**

$$\left\|\tilde{\Gamma}_Y(n_T)^{-1}\right\| = O_p(1) + O_p\left(\frac{kn_T}{T^{1/2}}\right) = O_p(1). \tag{A.13}$$

Now, we have

$$\begin{aligned}\mathsf{E}\|U_1(n_T)\| &= \mathsf{E}\left\|T^{-1}\sum_{t=1}^{T}\left[u_t(n_T) - u_t\right]Y_t(n_T)'\right\| \\ &\leq T^{-1}\sum_{t=1}^{T}\left\{\mathsf{E}\|u_t(n_T) - u_t\|^2\right\}^{1/2}\left\{\mathsf{E}\|Y_t(n_T)\|^2\right\}^{1/2},\end{aligned} \tag{A.14}$$

with

$$\mathsf{E}\left\|Y_t\left(n_T\right)\right\|^2 = 1 + n_T \mathrm{tr}\left[\Gamma_y\left(0\right)\right] \leq \left(1 + n_T\right)\delta \tag{A.15}$$

where $\delta = \max\left\{1, \mathrm{tr}\left[\Gamma_y\left(0\right)\right]\right\}$, and using (2.6),

$$\begin{aligned}
\mathsf{E}\left\|u_t\left(n_T\right) - u_t\right\|^2 &= \mathsf{E}\left\| \sum_{\tau=n_T+1}^{\infty} \Pi_\tau y_{t-\tau}^a \right\|^2 \\
&= \sum_{\tau_1=n_T+1}^{\infty} \sum_{\tau_2=n_T+1}^{\infty} \mathrm{tr}\left[\Pi_{\tau_1}' \Pi_{\tau_2} \Gamma_{y^a}\left(\tau_2 - \tau_1\right)\right] \\
&\leq \sum_{\tau_1=n_T+1}^{\infty} \sum_{\tau_2=n_T+1}^{\infty} \left\|\Pi_{\tau_1}\right\| \left\|\Pi_{\tau_2}\right\| \left\|\Gamma_{y^a}\left(\tau_2 - \tau_1\right)\right\| \\
&\leq \frac{C^2}{1-\rho^2} \left\|\Sigma_u\right\| \sum_{\tau_1=n_T+1}^{\infty} \sum_{\tau_2=n_T+1}^{\infty} \rho^{\left|\tau_2 - \tau_1\right|} \left\|\Pi_{\tau_1}\right\| \left\|\Pi_{\tau_2}\right\| \\
&\leq \frac{C^2}{1-\rho^2} \left\|\Sigma_u\right\| \left(\sum_{\tau=n_T+1}^{\infty} \left\|\Pi_\tau\right\| \right)^2
\end{aligned} \tag{A.16}$$

where C is a positive constant and $\Gamma_{y^a}\left(s-t\right) = \mathsf{E}\left[y_t^a y_s^{a\prime}\right]$ with $y_t^a = y_t - \mu_y = \sum_{v=0}^{\infty} \Psi_v u_{t-v}$, hence

$$\begin{aligned}
\mathsf{E}\left\|U_1\left(n_T\right)\right\| &\leq \frac{C}{\left(1-\rho^2\right)^{1/2}} \left\|\Sigma_u\right\|^{1/2} \left(1 + n_T\right)^{1/2} \delta^{1/2} \left(\sum_{\tau=n_T+1}^{\infty} \left\|\Pi_\tau\right\| \right) \\
&= C_1 \left(\frac{1+n_T}{n_T} \right)^{1/2} \left(n_T^{1/2} \sum_{\tau=n_T+1}^{\infty} \left\|\Pi_\tau\right\| \right) \\
&= O\left(1\right) \left(n_T^{1/2} \sum_{\tau=n_T+1}^{\infty} \left\|\Pi_\tau\right\| \right)
\end{aligned} \tag{A.17}$$

where C_1 is a positive constant. Then, we get

$$\left\|U_1\left(n_T\right)\right\| = O_p\left(1\right) \left(n_T^{1/2} \sum_{\tau=n_T+1}^{\infty} \left\|\Pi_\tau\right\| \right). \tag{A.18}$$

Since u_t and $Y_t\left(n_T\right)$ are independent, we have

$$\begin{aligned}
\mathsf{E}\left\|U_2\left(n_T\right)\right\|^2 &= \frac{1}{T^2} \sum_{t=1}^{T} \mathsf{E}\left[u_t' u_t\right] \mathsf{E}\left[Y_t\left(n_T\right)' Y_t\left(n_T\right)\right] \\
&= \frac{1}{T} \mathrm{tr}\left[\Sigma_u\right] \left(1 + n_T \mathrm{tr}\left[\Gamma_y\left(0\right)\right]\right) = O\left(\frac{k + k^2 n_T}{T} \right) = O\left(\frac{k^2 n_T}{T} \right),
\end{aligned} \tag{A.19}$$

and

$$\left\| U_2\left(n_T\right) \right\| = O_p\left(\frac{k n_T^{1/2}}{T^{1/2}}\right). \tag{A.20}$$

Then, by Assumption **2.4**, we get from (A.12), (A.13), (A.18) and (A.20),

$$\left\| \tilde{\Pi}\left(n_T\right) - \Pi\left(n_T\right) \right\| = o_p\left(1\right). \tag{A.21}$$

Finally, if Assumption **2.6** holds, we have

$$\left\| U_1\left(n_T\right) \right\| = O_p\left(\frac{n_T^{1/2}}{T^{1/2}}\right)\left(T^{1/2} \sum_{\tau=n_T+1}^{\infty} \left\| \Pi_\tau \right\|\right) = o_p\left(\frac{n_T^{1/2}}{T^{1/2}}\right) \tag{A.22}$$

and

$$\left\| \tilde{\Pi}\left(n_T\right) - \Pi\left(n_T\right) \right\| = O_p\left(\frac{k n_T^{1/2}}{T^{1/2}}\right). \tag{A.23}$$

□

PROOF OF PROPOSITION **3.2** Note first that

$$\begin{aligned}
\left\| \tilde{S}_Y\left(n_T\right) - S_Y\left(n_T\right) \right\| &= \left\| T^{1/2} l\left(n_T\right)' \operatorname{vec}\left[\tilde{\Omega}_Y\left(n_T\right) \tilde{\Gamma}_Y\left(n_T\right)^{-1} - \Omega_Y\left(n_T\right) \Gamma_Y\left(n_T\right)^{-1} \right] \right\| \\
&\leq T^{1/2} \left\| l\left(n_T\right) \right\| \left\| \operatorname{vec}\left[\tilde{\Omega}_Y\left(n_T\right) \tilde{\Gamma}_Y\left(n_T\right)^{-1} - \Omega_Y\left(n_T\right) \Gamma_Y\left(n_T\right)^{-1} \right] \right\| \\
&= \left\| l\left(n_T\right) \right\| \left\| T^{1/2} \left\{ \tilde{\Omega}_Y\left(n_T\right) \tilde{\Gamma}_Y\left(n_T\right)^{-1} - \Omega_Y\left(n_T\right) \Gamma_Y\left(n_T\right)^{-1} \right\} \right\| \\
&\leq M_2^{1/2} \left\{ \left\| q_1 + q_2 \right\| \right\} \\
&\leq M_2^{1/2} \left\{ \left\| q_1 \right\| + \left\| q_2 \right\| \right\}
\end{aligned} \tag{A.24}$$

where

$$q_1 = T^{1/2} \left[U_1\left(n_T\right) + U_2\left(n_T\right) \right] \left[\tilde{\Gamma}_Y\left(n_T\right)^{-1} - \Gamma_Y\left(n_T\right)^{-1} \right], \tag{A.25}$$

$$q_2 = T^{-1/2} \sum_{t=1}^{T} \left[u_t\left(n_T\right) - u_t \right] Y_t\left(n_T\right)' \Gamma_Y\left(n_T\right)^{-1}, \tag{A.26}$$

where $U_1(n_T)$ and $U_2(n_T)$ are defined in (A.11), with

$$\begin{aligned}\|q_1\| &= \left\|T^{1/2}\left[U_1(n_T)+U_2(n_T)\right]\left[\tilde{\Gamma}_Y(n_T)^{-1}-\Gamma_Y(n_T)^{-1}\right]\right\| \\ &\leq T^{1/2}\left\{\|U_1(n_T)\|+\|U_2(n_T)\|\right\}\left\|\tilde{\Gamma}_Y(n_T)^{-1}-\Gamma_Y(n_T)^{-1}\right\|_1, \end{aligned} \tag{A.27}$$

and

$$\begin{aligned}\mathsf{E}\|q_2\| &= \mathsf{E}\left\|T^{-1/2}\sum_{t=1}^{T}\left[u_t(n_T)-u_t\right]Y_t(n_T)'\Gamma_Y(n_T)^{-1}\right\| \\ &\leq T^{-1/2}\left\{\sum_{t=1}^{T}\mathsf{E}\left\|\left[u_t(n_T)-u_t\right]Y_t(n_T)'\Gamma_Y(n_T)^{-1}\right\|\right\} \\ &\leq T^{-1/2}\left\{\sum_{t=1}^{T}\mathsf{E}\|u_t(n_T)-u_t\|^2\right\}^{1/2}\left\{\sum_{t=1}^{T}\mathsf{E}\left\|Y_t(n_T)'\Gamma_Y(n_T)^{-1}\right\|^2\right\}^{1/2}. \end{aligned} \tag{A.28}$$

By Proposition **3.1**, (A.18) and (A.20), we can easily see under the Assumption **2.6** that

$$\begin{aligned}\|q_1\| &= T^{1/2}\left\{O_p(1)\left(n_T^{1/2}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\right)+O_p\left(\frac{n_T^{1/2}}{T^{1/2}}\right)\right\}O_p\left(\frac{n_T}{T^{1/2}}\right) \\ &= \left\{O_p(1)\left(T^{1/2}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\right)+O_p(1)\right\}O_p\left(\frac{n_T^{3/2}}{T^{1/2}}\right) \\ &= O_p\left(\frac{n_T^{3/2}}{T^{1/2}}\right). \end{aligned} \tag{A.29}$$

Moreover, we have

$$\begin{aligned}\mathsf{E}\left\|Y_t(n_T)'\Gamma_Y(n_T)^{-1}\right\|^2 &= \mathsf{E}\left\{\mathrm{tr}\left[\Gamma_Y(n_T)^{-1}Y_t(n_T)Y_t(n_T)'\Gamma_Y(n_T)^{-1}\right]\right\} \\ &= \mathrm{tr}\left[\Gamma_Y(n_T)^{-1}\mathsf{E}\left\{Y_t(n_T)Y_t(n_T)'\right\}\Gamma_Y(n_T)^{-1}\right] \\ &= \mathrm{tr}\left[\Gamma_Y(n_T)^{-1}\right]=O(1). \end{aligned} \tag{A.30}$$

Then, using (A.16), we get

$$\begin{aligned}\mathsf{E}\|q_2\| &\leq T^{-1/2}\left\{\sum_{t=1}^{T}\mathsf{E}\|u_t(n_T)-u_t\|^2\right\}^{1/2}\left\{\sum_{t=1}^{T}\mathsf{E}\left\|Y_t(n_T)'\Gamma_Y(n_T)^{-1}\right\|^2\right\}^{1/2}. \\ &= T^{-1/2}\left\{O(T)\left(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\right)^2\right\}^{1/2}\left\{O(T)\right\}^{1/2}\end{aligned}$$

$$= O(1)\left(T^{1/2}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\right) \tag{A.31}$$

and

$$\|q_2\| = O_p(1)\left(T^{1/2}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\right). \tag{A.32}$$

Hence

$$\left\|\tilde{S}_Y(n_T) - S_Y(n_T)\right\| = O_p\left(\frac{n_T^{3/2}}{T^{1/2}}\right) + O_p(1)\left(T^{1/2}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\right). \tag{A.33}$$

Therefore, under Assumptions **2.5** and **2.6**, we have

$$\left\|\tilde{S}_Y(n_T) - S_Y(n_T)\right\| = o_p(1). \tag{A.34}$$

Further, we can easily see that

$$\left(\frac{T^{1/2}}{n_T^{3/2}}\right)\left\|\tilde{S}_Y(n_T) - S_Y(n_T)\right\| = O_p(1) + O_p(1)\left(Tn_T^{-3/2}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\right) \tag{A.35}$$

where, by Assumption **2.5**,

$$Tn_T^{-3/2} \le c^{-3/2}T^{1-3\delta_2/2}, \tag{A.36}$$

since $n_T \ge cT^{\delta_2}$. Setting also

$$\delta_3 = 1 - \frac{3}{2}\delta_2, \tag{A.37}$$

we see that

$$\frac{1}{2} < \delta_3 < 1 \tag{A.38}$$

since $0 < \delta_2 < 1/3$. Hence, we get

$$\left(\frac{T^{1/2}}{n_T^{3/2}}\right)\left\|\tilde{S}_Y(n_T) - S_Y(n_T)\right\| = O_p(1) + O_p(1)\left(T^{\delta_3}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\right) \tag{A.39}$$

If, in addition, Assumption **2.7** holds, we get

$$\left(\frac{T^{1/2}}{n_T^{3/2}}\right)\left\|\tilde{S}_Y\left(n_T\right)-S_Y\left(n_T\right)\right\|=O_p\left(1\right)+o_p\left(1\right)=O_p\left(1\right), \tag{A.40}$$

and finally

$$\left\|\tilde{S}_Y\left(n_T\right)-S_Y\left(n_T\right)\right\|=O_p\left(\frac{n_T^{3/2}}{T^{1/2}}\right). \tag{A.41}$$

□

PROOF OF THEOREM **3.2** By the standard central limit theorem for stationary processes [see Anderson (1971, Section 7.7), Scott (1973, Theorem 2), Berk (1974, page 491), Lewis and Reinsel (1985, Theorem 3), Chung (2001, Theorem 9.1.5)] and under the assumption of independence between u_t and $Y_t\left(n_T\right)$ we have :

$$\frac{S_Y\left(n_T\right)}{\left\{l\left(n_T\right)'Q_Y\left(n_T\right)l\left(n_T\right)\right\}^{1/2}}\underset{T\to\infty}{\longrightarrow}N\left[0,1\right] \tag{A.42}$$

where $Q_Y\left(n_T\right)=\Gamma_Y\left(n_T\right)^{-1}\otimes\Sigma_u$ and $\Gamma_Y\left(n_T\right)=\mathsf{E}\left[Y_t\left(n_T\right)Y_t\left(n_T\right)'\right]$. Therefore, by Proposition **3.2** and Assumption **2.5** we finally conclude that

$$\frac{T^{1/2}l\left(n_T\right)'\mathrm{vec}\left[\tilde{\Pi}\left(n_T\right)-\Pi\left(n_T\right)\right]}{\left\{l\left(n_T\right)'Q_Y\left(n_T\right)l\left(n_T\right)\right\}^{1/2}}\underset{T\to\infty}{\longrightarrow}N\left[0,1\right]. \tag{A.43}$$

□

PROOF OF PROPOSITION **3.3** Let $\Sigma_u\left(T\right)=T^{-1}\sum_{t=1}^{T}u_tu_t'$. Then, by the triangular inequality, we have

$$\left\|\tilde{\Sigma}_u\left(n_T\right)-\Sigma_u\right\|\leq\left\|\tilde{\Sigma}_u\left(n_T\right)-\Sigma_u\left(T\right)\right\|+\left\|\Sigma_u\left(T\right)-\Sigma_u\right\| \tag{A.44}$$

where $\left\|\Sigma_u\left(T\right)-\Sigma_u\right\|=O_p\left(k/T^{1/2}\right)$ and

$$\left\|\tilde{\Sigma}_u\left(n_T\right)-\Sigma_u\left(T\right)\right\| \quad\leq\quad \frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{u}_t\left(n_T\right)\tilde{u}_t\left(n_T\right)'-u_tu_t'\right\|$$

$$\leq \frac{1}{T}\sum_{t=1}^{T}\Big\{\big\|\tilde{u}_t(n_T)-u_t\big\|\big\|\tilde{u}_t(n_T)\big\|+\big\|u_t\big\|\big\|\tilde{u}_t(n_T)-u_t\big\|\Big\}. \tag{A.45}$$

Moreover, we have

$$\begin{aligned}\big\|\tilde{u}_t(n_T)-u_t\big\|^2 &\leq 2\big\|\tilde{u}_t(n_T)-u_t(n_T)\big\|^2+2\big\|u_t(n_T)-u_t\big\|^2\\ &\leq 2\Big\|\tilde{\Pi}(n_T)-\Pi(n_T)\Big\|^2\big\|Y_t(n_T)\big\|^2+2\Big\|\sum_{\tau=n_T+1}^{\infty}\Pi_\tau y_{t-\tau}^a\Big\|^2\end{aligned} \tag{A.46}$$

where $\big\|\tilde{\Pi}(n_T)-\Pi(n_T)\big\|^2=O_p\left(k^2n_T/T\right)$, $\mathsf{E}\big\|Y_t(n_T)\big\|^2=O_p\left(kn_T\right)$ and

$$\mathsf{E}\Big\|\sum_{\tau=n_T+1}^{\infty}\Pi_\tau y_{t-\tau}^a\Big\|^2=O_p(k)\Big(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\Big)^2=O_p\Big(k\rho^{2n_T}\Big), \tag{A.47}$$

hence

$$\big\|\tilde{u}_t(n_T)-u_t\big\|^2=O_p\Big(\frac{k^3n_T^2}{T}\Big) \tag{A.48}$$

and

$$\big\|\tilde{u}_t(n_T)-u_t\big\|=O_p\Big(\frac{k^{3/2}n_T}{T^{1/2}}\Big). \tag{A.49}$$

Finally, we get

$$\Big\|\tilde{\Sigma}_u(n_T)-\Sigma_u(T)\Big\|=O_p\Big(\frac{k^2n_T}{T^{1/2}}\Big), \tag{A.50}$$

since $\mathsf{E}\big\|u_t\big\|^2=O_p(k)$. Hence

$$\Big\|\tilde{\Sigma}_u(n_T)-\Sigma_u\Big\|=O_p\Big(\frac{k^2n_T}{T^{1/2}}\Big) \tag{A.51}$$

and, similarly, as in the proof of Proposition **3.1**, it is easy to see that

$$\Big\|\tilde{\Sigma}_u(n_T)^{-1}-\Sigma_u^{-1}\Big\|=O_p\Big(\frac{k^2n_T}{T^{1/2}}\Big). \tag{A.52}$$

□

PROOF OF PROPOSITION **3.4** Let

$$\Gamma_X(T) = T^{-1}\sum_{t=1}^{T} X_t X_t', \quad \Upsilon_X(T) = \Gamma_X(T) \otimes \Sigma_u^{-1}, \quad Q_X(T) = \left\{R'\Upsilon_X(T)R\right\}^{-1}. \tag{A.53}$$

Then

$$\begin{aligned}\left\|\tilde{Q}_X(n_T)^{-1} - Q_X^{-1}\right\|_1 &\leq \left\|\tilde{Q}_X(n_T)^{-1} - Q_X^{-1}\right\| \\ &\leq \left\|\tilde{Q}_X(n_T)^{-1} - Q_X(T)^{-1}\right\| + \left\|Q_X(T)^{-1} - Q_X^{-1}\right\|,\end{aligned} \tag{A.54}$$

with

$$\left\|Q_X(T)^{-1} - Q_X^{-1}\right\| \leq \|R\|^2\|\Upsilon_X(T) - \Upsilon_X\| = \|R\|^2\|\Sigma_u^{-1}\|\|\Gamma_X(T) - \Gamma_X\| \tag{A.55}$$

where $\|R\|^2 = r_{\bar{p}}$, $\|\Sigma_u^{-1}\| = O_p(1)$ and

$$\|\Gamma_X(T) - \Gamma_X\| = O_p\left(\frac{hk}{T^{1/2}}\right). \tag{A.56}$$

Hence

$$\left\|Q_X(T)^{-1} - Q_X^{-1}\right\| = O_p\left(\frac{r_{\bar{p}}hk}{T^{1/2}}\right). \tag{A.57}$$

Moreover, we have

$$\left\|\tilde{Q}_X(n_T)^{-1} - Q_X(T)^{-1}\right\| \leq \|R\|^2\left\|\tilde{\Upsilon}_X(n_T) - \Upsilon_X(T)\right\| \tag{A.58}$$

where

$$\left\|\tilde{\Upsilon}_X(n_T) - \Upsilon_X(T)\right\| \leq \left\|\tilde{\Gamma}_X(n_T)\right\|\left\|\tilde{\Sigma}_u(n_T)^{-1} - \Sigma_u^{-1}\right\| + \left\|\tilde{\Gamma}_X(n_T) - \Gamma_X(T)\right\|\|\Sigma_u^{-1}\|, \tag{A.59}$$

$$\left\|\tilde{\Gamma}_X(n_T) - \Gamma_X(T)\right\| \leq \frac{1}{T}\sum_{t=1}^{T}\left\{\left\|\tilde{X}_t(n_T)\right\|\left\|\tilde{X}_t(n_T) - X_t\right\| + \left\|\tilde{X}_t(n_T) - X_t\right\|\|X_t\|\right\}, \tag{A.60}$$

with

$$\left\|\tilde{X}_t(n_T) - X_t\right\| = \left\{\sum_{j=0}^{\bar{p}}\left\|\tilde{u}_{t-j}(n_T) - u_{t-j}\right\|^2\right\}^{1/2}$$

$$= \left(\bar{p}+1\right)^{1/2}\left\|\tilde{u}_t\left(n_T\right)-u_t\right\|$$

$$\leq h^{1/2}\left\|\tilde{u}_t\left(n_T\right)-u_t\right\| = O_p\left(\frac{h^{1/2}k^{3/2}n_T}{T^{1/2}}\right), \tag{A.61}$$

using (A.49). Further, since $\mathsf{E}\left\|X_t\right\|^2 = O_p\left(hk\right)$ and $\left\|\Sigma_u^{-1}\right\| = O_p\left(1\right)$, we have

$$\left\|\tilde{\Gamma}_X\left(n_T\right)-\Gamma_X\left(T\right)\right\| = O_p\left(\frac{hk^2n_T}{T^{1/2}}\right), \quad \left\|\tilde{\Upsilon}_X\left(n_T\right)-\Upsilon_X\left(T\right)\right\| = O_p\left(\frac{hk^3n_T}{T^{1/2}}\right), \tag{A.62}$$

and then

$$\left\|\tilde{Q}_X\left(n_T\right)^{-1}-Q_X\left(T\right)^{-1}\right\| = O_p\left(\frac{r_{\bar{p}}hk^3n_T}{T^{1/2}}\right). \tag{A.63}$$

Hence

$$\left\|\tilde{Q}_X\left(n_T\right)^{-1}-Q_X^{-1}\right\| = O_p\left(\frac{r_{\bar{p}}hk^3n_T}{T^{1/2}}\right), \quad \left\|\tilde{Q}_X\left(n_T\right)^{-1}-Q_X^{-1}\right\|_1 = O_p\left(\frac{r_{\bar{p}}hk^3n_T}{T^{1/2}}\right) \tag{A.64}$$

and, as in proof of Proposition **3.1**, we finally get

$$\left\|\tilde{Q}_X\left(n_T\right)-Q_X\right\|_1 = O_p\left(\frac{r_{\bar{p}}hk^3n_T}{T^{1/2}}\right). \tag{A.65}$$

□

PROOF OF THEOREM **3.3** Recall that $\tilde{\eta}-\eta = \tilde{Q}_X\left(n_T\right)\tilde{\Omega}_X\left(n_T\right)$. Then

$$\left\|\tilde{\eta}-\eta\right\| \leq \left\|\tilde{Q}_X\left(n_T\right)\right\|_1\left\|\tilde{\Omega}_X\left(n_T\right)-\Omega_X\right\| + \left\|\tilde{Q}_X\left(n_T\right)-Q_X\right\|_1\left\|\Omega_X\right\| + \left\|Q_X\right\|_1\left\|\Omega_X\right\| \tag{A.66}$$

where, by Proposition **3.4**,

$$\left\|\tilde{Q}_X\left(n_T\right)-Q_X\right\|_1 = O_p\left(\frac{r_{\bar{p}}hk^3n_T}{T^{1/2}}\right), \quad \left\|\tilde{Q}_X\left(n_T\right)\right\|_1 = O_p\left(1\right). \tag{A.67}$$

Let also

$$W_X = \frac{1}{T}\sum_{t=1}^{T}u_tX_t'. \tag{A.68}$$

Then one easily sees that

$$\Omega_X = R' \mathrm{vec}\left[\Sigma_u^{-1} W_X\right] \tag{A.69}$$

and

$$\mathsf{E}\|\Omega_X\|^2 \leq \|R\|^2 \|\Sigma_u^{-1}\|^2 \mathsf{E}\|W_X\|^2 \tag{A.70}$$

where, by independence between u_t and X_t

$$\mathsf{E}\|W_X\|^2 = O_p\left(\frac{hk^2}{T}\right). \tag{A.71}$$

Hence

$$\|\Omega_X\| = O_p\left(\frac{r_{\bar{p}}^{1/2} h^{1/2} k}{T^{1/2}}\right). \tag{A.72}$$

Now, consider the term $\|\tilde{\Omega}_X(n_T) - \Omega_X\|$. It can be easily seen that

$$\begin{aligned} \left\|\tilde{\Omega}_X(n_T) - \Omega_X\right\| \quad \leq \quad & \|R\| \Big[\left\|\tilde{\Sigma}_u(n_T)^{-1}\right\| \left\{ \left\|\tilde{W}_X^1(n_T)\right\| + \left\|\tilde{W}_X^2(n_T)\right\| + \left\|\tilde{W}_X^3(n_T)\right\| \right\} \\ & + \left\|\tilde{\Sigma}_u(n_T)^{-1} - \Sigma_u^{-1}\right\| \|W_X\| \Big], \end{aligned} \tag{A.73}$$

where

$$\tilde{W}_X^1(n_T) = \frac{1}{T} \sum_{t=1}^{T} \left[e_t(n_T) - u_t\right] \left[\tilde{X}_t(n_T) - X_t\right]', \tag{A.74}$$

$$\tilde{W}_X^2(n_T) = \frac{1}{T} \sum_{t=1}^{T} \left[e_t(n_T) - u_t\right] X_t', \quad \tilde{W}_X^3(n_T) = \frac{1}{T} \sum_{t=1}^{T} u_t \left[\tilde{X}_t(n_T) - X_t\right]'. \tag{A.75}$$

By Proposition **3.3** we have

$$\left\|\tilde{\Sigma}_u(n_T)^{-1}\right\| = O_p(1) + O_p\left(\frac{k^2 n_T}{T^{1/2}}\right), \quad \left\|\tilde{\Sigma}_u(n_T)^{-1} - \Sigma_u^{-1}\right\| = O_p\left(\frac{k^2 n_T}{T^{1/2}}\right). \tag{A.76}$$

Moreover, using (2.17), (3.20) and (A.61), one easily sees that

$$\|e_t(n_T) - u_t\| \leq \left\|\tilde{X}_t(n_T) - X_t\right\| \|R\| \|\eta\| = O_p\left(\frac{r_{\bar{p}} h^{1/2} k^{3/2} n_T}{T^{1/2}}\right). \tag{A.77}$$

Hence

$$\left\|\tilde{W}_X^1(n_T)\right\| \leq \left\{\frac{1}{T}\sum_{t=1}^{T}\left\|e_t(n_T)-u_t\right\|^2\right\}^{1/2}\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{X}_t(n_T)-X_t\right\|^2\right\}^{1/2} = O_p\left(\frac{r_{\bar{p}}hk^3n_T^2}{T}\right). \tag{A.78}$$

Further, setting $F = \left[\mu_\Phi, I_k - \Phi_0, \Phi_1, \ldots, \Phi_{\bar{p}}, \Theta_1, \ldots, \Theta_{\bar{p}}\right]$, one easily sees that

$$\left\|\tilde{W}_X^2(n_T)\right\|^2 \leq \left\|F\right\|^2\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{X}_t(n_T)-X_t\right]X_t'\right\|^2, \tag{A.79}$$

with

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{X}_t(n_T)-X_t\right]X_t'\right\|^2 = \sum_{j=0}^{\bar{p}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{u}_{t-j}(n_T)-u_{t-j}\right]X_t'\right\|^2 \tag{A.80}$$

and

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\left[u_{t-j}-\tilde{u}_{t-j}(n_T)\right]X_t'\right\|^2 \leq 2\left\|\tilde{W}_X^{21}(n_T)\right\|^2 + 2\left\|\tilde{W}_X^{22}(n_T)\right\|^2 \tag{A.81}$$

where $\tilde{W}_X^{21}(n_T) = T^{-1}\sum_{t=1}^{T}\left[\tilde{u}_{t-j}(n_T)-u_{t-j}(n_T)\right]X_t'$ and $\tilde{W}_X^{22}(n_T) = T^{-1}\sum_{t=1}^{T}\left[u_{t-j}(n_T)-u_{t-j}\right]X_t'$.

In particular,

$$\left\|\tilde{W}_X^{21}(n_T)\right\|^2 \leq \left\|\Pi(n_T)-\tilde{\Pi}(n_T)\right\|^2\left\|W_X^Y(n_T)\right\|^2 \tag{A.82}$$

where

$$\left\|W_X^Y(n_T)\right\|^2 = \left\|\frac{1}{T}\sum_{t=1}^{T}Y_{t-j}(n_T)X_t'\right\|^2 = \left\|\frac{1}{T}\sum_{t=1}^{T}X_t\right\|^2 + \sum_{\tau=1}^{n_T}\left\|\frac{1}{T}\sum_{t=1}^{T}y_{t-j-\tau}X_t'\right\|^2, \tag{A.83}$$

and given the VARMA structure of y_t as described above, one easily sees that

$$\left\|\frac{1}{T}\sum_{t=1}^{T}X_t\right\|^2 = O_p\left(\frac{hk}{T}\right) \tag{A.84}$$

and

$$\mathsf{E}\left\|\frac{1}{T}\sum_{t=1}^{T}y_{t-j-\tau}X_t'\right\|^2 \leq \frac{hk^2C_1\rho_1^{\tau+j}}{T} \tag{A.85}$$

for some constants $C_1 > 0$ and $0 < \rho < \rho_1 < 1$. Consequently, we get

$$\sum_{\tau=1}^{n_T}\mathsf{E}\left\|\frac{1}{T}\sum_{t=1}^{T}y_{t-j-\tau}X_t'\right\|^2 \leq \sum_{\tau=1}^{n_T}\frac{C_1\rho_1^{\tau+j}}{T} = \frac{hk^2C_1\rho_1^j}{T(1-\rho_1)} = \frac{hk^2C_2}{T}, \tag{A.86}$$

with $C_2 = C_1\rho_1^j/\ (1-\rho_1)$, and then

$$\left\| W_X^Y\left(n_T\right)\right\|^2 = O_p\left(\frac{hk^2}{T}\right). \tag{A.87}$$

Hence, using (A.23) and (A.87), we easily show that

$$\left\| \tilde{W}_X^{21}\left(n_T\right)\right\|^2 = O_p\left(\frac{hk^4 n_T}{T^2}\right). \tag{A.88}$$

In addition, we have

$$\begin{aligned}
\mathsf{E}\left\|\tilde{W}_X^{22}\left(n_T\right)\right\| &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\mathsf{E}\left\|y_{t-j-\tau}^{a} X_t'\right\| \\
&\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\left\{\mathsf{E}\left\|y_{t-j-\tau}^{a}\right\|^2\right\}^{1/2}\left\{\mathsf{E}\left\|X_t\right\|^2\right\}^{1/2} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\{\mathrm{tr}\left[\Gamma_{y^a}\left(0\right)\right]\right\}^{1/2}\left\{\mathrm{tr}\left[\Gamma_X\left(0\right)\right]\right\}^{1/2}\left(\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right) \\
&= O_p\left(\rho^{n_T}\right)
\end{aligned} \tag{A.89}$$

where $\Gamma_X\left(0\right) = \Gamma_X$. Therefore

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{u}_{t-j}\left(n_T\right) - u_{t-j}\right]X_t'\right\|^2 = O_p\left(\frac{hk^4 n_T}{T^2}\right), \tag{A.90}$$

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{X}_t\left(n_T\right) - X_t\right]X_t'\right\|^2 = O_p\left(\frac{h^2k^4 n_T}{T^2}\right), \tag{A.91}$$

and then

$$\left\|\tilde{W}_X^{2}\left(n_T\right)\right\| = O_p\left(\frac{hk^2 n_T^{1/2}}{T}\right). \tag{A.92}$$

Furthermore, one can easily see that

$$\left\|\tilde{W}_X^{3}\left(n_T\right)\right\|^2 = \sum_{j=0}^{\bar{p}}\left\|\frac{1}{T}\sum_{t=1}^{T}u_t\left[\tilde{u}_{t-j}\left(n_T\right) - u_{t-j}\right]'\right\|^2 \tag{A.93}$$

where

$$\left\| \frac{1}{T} \sum_{t=1}^{T} u_t \left[\tilde{u}_{t-j}\left(n_T\right) - u_{t-j} \right]' \right\| \leq \left\| \tilde{W}_X^{31}\left(n_T\right) \right\| + \left\| \tilde{W}_X^{32}\left(n_T\right) \right\|, \tag{A.94}$$

with $\tilde{W}_X^{31}\left(n_T\right) = T^{-1} \sum_{t=1}^{T} u_t \left[\tilde{u}_{t-j}\left(n_T\right) - u_{t-j}\left(n_T\right) \right]'$ and $\tilde{W}_X^{32}\left(n_T\right) = T^{-1} \sum_{t=1}^{T} u_t \left[u_{t-j}\left(n_T\right) - u_{t-j} \right]'$.

More especially, we have

$$\left\| \tilde{W}_X^{31}\left(n_T\right) \right\| \leq \left\| W_u^Y\left(n_T\right) \right\| \left\| \Pi\left(n_T\right) - \tilde{\Pi}\left(n_T\right) \right\|. \tag{A.95}$$

Therefore, by independence between u_t and $Y_{t-j}\left(n_T\right)$ for $j \geq 0$,

$$\left\| W_u^Y\left(n_T\right) \right\| = \left\| \frac{1}{T} \sum_{t=1}^{T} u_t Y_{t-j}\left(n_T\right)' \right\| = O_p\left(\frac{k n_T^{1/2}}{T^{1/2}} \right), \tag{A.96}$$

and in view of (A.23), we easily get

$$\left\| \tilde{W}_X^{31}\left(n_T\right) \right\| = O_p\left(\frac{k^2 n_T}{T} \right). \tag{A.97}$$

In the other hand, we have

$$\begin{aligned}
\mathsf{E}\left\| \tilde{W}_X^{32}\left(n_T\right) \right\| &\leq \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=n_T+1}^{\infty} \mathsf{E}\left\| u_t y_{t-j-\tau}^{a\,\prime} \right\| \left\| \Pi_\tau \right\| \\
&\leq \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=n_T+1}^{\infty} \left\{ \mathsf{E}\left\| y_{t-j-\tau}^{a} \right\|^2 \right\}^{1/2} \left\{ \mathsf{E}\left\| u_t \right\|^2 \right\}^{1/2} \left\| \Pi_\tau \right\| \\
&= \frac{1}{T} \sum_{t=1}^{T} \left\{ \mathrm{tr}\left[\Gamma_{y^a}\left(0\right) \right] \right\}^{1/2} \left\{ \mathrm{tr}\left[\Sigma_u \right] \right\}^{1/2} \left(\sum_{\tau=n_T+1}^{\infty} \left\| \Pi_\tau \right\| \right) \\
&= O_p\left(\rho^{n_T} \right).
\end{aligned} \tag{A.98}$$

Therefore, it follows that

$$\left\| \frac{1}{T} \sum_{t=1}^{T} u_t \left[\tilde{u}_{t-j}\left(n_T\right) - u_{t-j} \right]' \right\|^2 = O_p\left(\frac{k^4 n_T^2}{T^2} \right), \tag{A.99}$$

$$\left\| \frac{1}{T} \sum_{t=1}^{T} u_t \left[\tilde{X}_t\left(n_T\right) - X_t \right]' \right\|^2 = O_p\left(\frac{h k^4 n_T^2}{T^2} \right), \tag{A.100}$$

and then

$$\left\| \tilde{W}_X^{3}\left(n_T\right) \right\| = O_p\left(\frac{h^{1/2} k^2 n_T}{T} \right). \tag{A.101}$$

Finally, one easily sees that

$$\begin{aligned}\left\|\tilde{\Omega}_X\left(n_T\right)-\Omega_X\right\| &\leq \|R\|\left\|\tilde{\Sigma}_u\left(n_T\right)^{-1}\right\|\left\{\left\|\tilde{W}_X^1\left(n_T\right)\right\|+\left\|\tilde{W}_X^2\left(n_T\right)\right\|+\left\|\tilde{W}_X^3\left(n_T\right)\right\|\right\} \\ &\quad +\|R\|\left\|\tilde{\Sigma}_u\left(n_T\right)^{-1}-\Sigma_u^{-1}\right\|\|W_X\| \\ &= O_p\left(\frac{r_{\bar{p}}hk^3n_T^2}{T}\right). \end{aligned} \tag{A.102}$$

As a result, we get

$$\left\|\tilde{\eta}-\eta\right\| = O_p\left(\frac{r_{\bar{p}}^{1/2}h^{1/2}k}{T^{1/2}}\right)+O_p\left(\frac{r_{\bar{p}}hk^3n_T^2}{T}\right), \tag{A.103}$$

or equivalently

$$\left\|\tilde{\eta}-\eta\right\| = O_p\left(\frac{1}{T^{1/2}}\right)+O_p\left(\frac{n_T^2}{T}\right). \tag{A.104}$$

Furthermore, in view of Assumption **2.3**,

$$\left\|\tilde{\eta}-\eta\right\| = o_p\left(1\right). \tag{A.105}$$

Moreover, one sees that

$$\left\|T^{1/2}\left(\tilde{\eta}-\eta\right)\right\| = O_p\left(1\right)+O_p\left(\frac{n_T^2}{T^{1/2}}\right). \tag{A.106}$$

Hence, under the Assumption **2.8**, we have

$$\left\|\tilde{\eta}-\eta\right\| = O_p\left(\frac{1}{T^{1/2}}\right), \tag{A.107}$$

or equivalently

$$\left\|\tilde{\eta}-\eta\right\| = O_p\left(\frac{r_{\bar{p}}^{1/2}h^{1/2}k}{T^{1/2}}\right). \tag{A.108}$$

□

PROOF OF PROPOSITION **3.5** Note that

$$\begin{aligned}\left\|\tilde{S}_X(n_T) - S_X\right\| &= T^{1/2}\left\|\tilde{Q}_X(n_T)\,\tilde{\Omega}_X(n_T) - Q_X\Omega_X\right\| \\ &\leq T^{1/2}\left\|\tilde{Q}_X\right\|_1\left\|\tilde{\Omega}_X(n_T) - \Omega_X\right\| + T^{1/2}\left\|\tilde{Q}_X(n_T) - Q_X\right\|_1\|\Omega_X\| \qquad \text{(A.109)}\end{aligned}$$

where by Proposition **3.4** and Theorem **3.3**, we have :

$$\left\|\tilde{Q}_X(n_T)\right\|_1 = O_p(1)\,, \quad \left\|\tilde{Q}_X(n_T) - Q_X\right\|_1 = O_p\left(\frac{r_{\bar{p}}hk^3n_T}{T^{1/2}}\right), \qquad \text{(A.110)}$$

$$\|\Omega_X\| = O_p\left(\frac{r_{\bar{p}}^{1/2}h^{1/2}k}{T^{1/2}}\right), \quad \left\|\tilde{\Omega}_X(n_T) - \Omega_X\right\| = O_p\left(\frac{r_{\bar{p}}hk^4n_T^2}{T}\right), \qquad \text{(A.111)}$$

and finally

$$\left\|\tilde{S}_X(n_T) - S_X\right\| = O_p\left(\frac{r_{\bar{p}}hkn_T^2}{T^{1/2}}\right). \qquad \text{(A.112)}$$

□

PROOF OF THEOREM **3.4** By the standard central limit theorem for stationary processes [see Anderson (1971, Section 7.7), Scott (1973, Theorem 2) and Chung (2001, Theorem 9.1.5)], and under the assumption of independence between u_t and X_t, we have :

$$T^{1/2}\Omega_X \underset{T\to\infty}{\longrightarrow} N\left[0,\, Q_X^{-1}\right]. \qquad \text{(A.113)}$$

Then

$$S_X = T^{1/2}Q_X\Omega_X \underset{T\to\infty}{\longrightarrow} N\left[0,\, Q_X\right] \qquad \text{(A.114)}$$

and by Proposition **3.5** and under the assumption **2.8** we conclude that

$$T^{1/2}\left(\tilde{\eta} - \eta\right) = \tilde{S}_X(n_T) \underset{T\to\infty}{\longrightarrow} N\left[0,\, Q_X\right]. \qquad \text{(A.115)}$$

□

PROOF OF PROPOSITION **3.6** By the triangular inequality we have

$$\left\|\tilde{\Sigma}_e(n_T) - \Sigma_u\right\| \leq \left\|\tilde{\Sigma}_e(n_T) - \Sigma_u(T)\right\| + \left\|\Sigma_u(T) - \Sigma_u\right\| \tag{A.116}$$

where

$$\begin{aligned}\left\|\tilde{\Sigma}_e(n_T) - \Sigma_u(T)\right\| &\leq \frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{e}_t(n_T)\,\tilde{e}_t(n_T)' - u_t u_t'\right\| \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\left\{\left\|\tilde{e}_t(n_T) - u_t\right\|\left\|\tilde{e}_t(n_T)\right\| + \left\|u_t\right\|\left\|\tilde{e}_t(n_T) - u_t\right\|\right\},\end{aligned} \tag{A.117}$$

and

$$\begin{aligned}\left\|\tilde{e}_t(n_T) - u_t\right\| &= \left\|\left[X_t' \otimes I_k\right]R\eta - \left[\tilde{X}_t(n_T)' \otimes I_k\right]R\tilde{\eta}\right\| \\ &\leq \left\|I_k\right\|\left\|R\right\|\left\{\left\|X_t\right\|\left\|\eta - \tilde{\eta}\right\| + \left\|X_t - \tilde{X}_t(n_T)\right\|\left\|\tilde{\eta}\right\|\right\} \\ &= O_p\left(\frac{r_{\bar{p}}hk^2}{T^{1/2}}\right) + O_p\left(\frac{r_{\bar{p}}h^{1/2}k^2n_T}{T^{1/2}}\right) = O_p\left(\frac{r_{\bar{p}}h^{1/2}k^2n_T}{T^{1/2}}\right).\end{aligned} \tag{A.118}$$

Therefore, we get

$$\left\|\tilde{\Sigma}_e(n_T) - \Sigma_u(T)\right\| = O_p\left(\frac{r_{\bar{p}}h^{1/2}k^{5/2}n_T}{T^{1/2}}\right), \tag{A.119}$$

and finally

$$\left\|\tilde{\Sigma}_e(n_T) - \Sigma_u\right\| = \left\|\tilde{\Sigma}_e(n_T)^{-1} - \Sigma_u^{-1}\right\| = O_p\left(\frac{r_{\bar{p}}h^{1/2}k^{5/2}n_T}{T^{1/2}}\right). \tag{A.120}$$

□

PROOF OF LEMMA **4.1** Under the invertibility condition of the echelon form VARMA representation we have $\det[\Theta(z)] \neq 0$, $|z| \leq 1$. Then there exists a positive constant ε, such that

$$\Theta(z)^{-1} = \sum_{\tau=0}^{\infty}\Lambda_\tau(\eta)\,z^\tau, \quad |z| < 1+\varepsilon. \tag{A.121}$$

Moreover, there exist real constants $(\varsigma,\zeta) > 0$ and $\tau \geq \tau_0$ $(\tau,\tau_0 \in \mathbb{Z})$, such that $\Lambda_\tau(\eta)(1+\varsigma)^\tau \to 0$ as

$\tau \to \infty$, and $\Lambda_{lc,\tau}(\eta) \leq \zeta(1+\varsigma)^{-\tau}$, $|z| < 1+\varsigma$, where $\Lambda_{lc,\tau}(\eta)$ is the component of $\Lambda_\tau(\eta)$ in the l-th row and c-th column $(l, c = 1, ..., k)$ and $0 < \varsigma < \varepsilon$. This means that all components of $\Lambda_\tau(\eta)$ are geometrically bounded. Further, let $\rho = (1+\varsigma)^{-1}$ so that $\rho \in (0,1)$, then $\|\Lambda_\tau(\eta)\| \leq C\rho^\tau$, with $C = k\zeta$. In particular, there exists a positive constant κ such that $1+\kappa^{-1} < \rho^{-1}$. Hence for $|z| \leq 1+\kappa^{-1}$

$$\begin{aligned}\sum_{\tau=0}^{\infty} \|\Lambda_\tau(\eta)\| |z|^\tau &\leq \sum_{\tau=0}^{\infty} \|\Lambda_\tau(\eta)\| (1+\kappa^{-1})^\tau \\ &\leq \sum_{\tau=0}^{\infty} C\rho^\tau (1+\kappa^{-1})^\tau \\ &= \sum_{\tau=0}^{\infty} C\left[\rho(1+\kappa^{-1})\right]^\tau = \frac{C\kappa}{\kappa - \rho(\kappa+1)} < \infty. \end{aligned} \tag{A.122}$$

Let also $\Lambda_{lc,\tau}(\breve{\eta})$ and $\Lambda_{lc,\tau}(\eta)$ be the components of $\Lambda_\tau(\breve{\eta})$ and $\Lambda_\tau(\eta)$, respectively. Then

$$\left|\Lambda_{lc,\tau}(\breve{\eta}) - \Lambda_{lc,\tau}(\eta)\right| = (\tau!)^{-1}\left|\left[\Lambda_{lc}^{(\tau)}(\breve{\eta})(z) - \Lambda_{lc}^{(\tau)}(\eta)(z)\right]\Big|_{z=0}\right| \tag{A.123}$$

where $|.|$ stands for the euclidean norm, and $\Lambda_{lc}^{(\tau)}(\breve{\eta})$ and $\Lambda_{lc}^{(\tau)}(\eta)$ designate the τ-th derivatives of $\Lambda_{lc}(\breve{\eta})$ and $\Lambda_{lc}(\eta)$ with respect to z, respectively. Hence, applying the Cauchy inequality to the derivatives of an analytic function, here $\Lambda(\eta)(z)$ [see Ahlfors (1966, Page 122), and Churchill and Brown (1990, Page 130)], we get

$$\left|\left[\Lambda_{lc}^{(\tau)}(\breve{\eta})(z) - \Lambda_{lc}^{(\tau)}(\eta)(z)\right]\Big|_{z=0}\right| \leq (\tau!)(1+\kappa^{-1})^{-\tau} \max_{|z|=1+\kappa^{-1}} \left|\Lambda_{lc}(\breve{\eta})(z) - \Lambda_{lc}(\eta)(z)\right|, \tag{A.124}$$

then

$$\begin{aligned}\left|\Lambda_{lc,\tau}(\breve{\eta}) - \Lambda_{lc,\tau}(\eta)\right| &\leq (1+\kappa^{-1})^{-\tau} \max_{|z|=1+\kappa^{-1}} \left|\Lambda_{lc}(\breve{\eta})(z) - \Lambda_{lc}(\eta)(z)\right| \\ &\leq (1+\kappa^{-1})^{-\tau} \max_{|z|=1+\kappa^{-1}} \left|\left[\det\{\breve{\Theta}_\tau(z)\}\right]^{-1}\breve{\theta}_{lc,\tau}^{+}(z) - \left[\det\{\Theta_\tau(z)\}\right]^{-1}\theta_{lc,\tau}^{+}(z)\right| \\ &\leq (1+\kappa^{-1})^{-\tau} \max_{|z|=1+\kappa^{-1}} \left|\left[\det\{\breve{\Theta}_\tau(z)\}\right]^{-1} - \left[\det\{\Theta_\tau(z)\}\right]^{-1}\right| \left|\breve{\theta}_{lc,\tau}^{+}(z)\right| \\ &\quad + (1+\kappa^{-1})^{-\tau} \max_{|z|=1+\kappa^{-1}} \left|\left[\det\{\Theta_\tau(z)\}\right]^{-1}\right| \left|\breve{\theta}_{lc,\tau}^{+}(z) - \theta_{lc,\tau}^{+}(z)\right|, \end{aligned} \tag{A.125}$$

for $\tau \in Z$ and $|z| \leq 1+\kappa^{-1}$, where the polynomials $\breve{\theta}_{lc,\tau}^{+}(z)$ and $\theta_{lc,\tau}^{+}(z)$ are the (l,c)-th components of

the adjoint matrices of $\breve{\Theta}(z)$ and $\Theta(z)$, respectively. By assumption $\|\breve{\eta}-\eta\| = O_p\left(T^{-1/2}\right)$, hence $\|\breve{\Theta}(z) - \Theta(z)\| = O_p\left(T^{-1/2}\right)$, $|z| \leq 1+\kappa^{-1}$. Consequently, we have

$$\left\| \left[\det\left\{\breve{\Theta}_\tau(z)\right\}\right]^{-1} - \left[\det\left\{\Theta_\tau(z)\right\}\right]^{-1} \right\| = O_p\left(T^{-1/2}\right), \tag{A.126}$$

$$\left| \breve{\theta}^{+}_{lc,\tau}(z) - \theta^{+}_{lc,\tau}(z) \right| = O_p\left(T^{-1/2}\right), \tag{A.127}$$

and then

$$\left| \Lambda_{lc,\tau}(\breve{\eta}) - \Lambda_{lc,\tau}(\eta) \right| \leq C\left(1+\kappa^{-1}\right)^{-\tau} T^{-1/2}. \tag{A.128}$$

Hence

$$\left\| \Lambda_\tau(\breve{\eta}) - \Lambda_\tau(\eta) \right\| \leq C\left(1+\kappa^{-1}\right)^{-\tau} T^{-1/2} \tag{A.129}$$

and finally

$$T^{1/2}\left(1+\kappa^{-1}\right)^{\tau} \left\| \Lambda_\tau(\breve{\eta}) - \Lambda_\tau(\eta) \right\| = O_p(1). \tag{A.130}$$

As a result, we get

$$\left\| \Lambda_\tau(\breve{\eta}) \right\| \leq C\left[\rho^\tau + \left(1+\kappa^{-1}\right)^{-\tau} T^{-1/2}\right]. \tag{A.131}$$

□

PROOF OF PROPOSITION **4.1** By the triangular inequality we have

$$\begin{aligned} \left\| \tilde{\Sigma}_u(\tilde{\eta}) - \Sigma_u \right\| &\leq \frac{1}{T}\sum_{t=1}^{T} \left\| u_t(\tilde{\eta})\, u_t(\tilde{\eta})' - u_t u_t' \right\| + O_p\left(T^{-1/2}\right) \\ &\leq \frac{1}{T}\sum_{t=1}^{T} \left\{ \left\| u_t(\tilde{\eta}) - u_t \right\| \left\| u_t(\tilde{\eta}) \right\| + \left\| u_t \right\| \left\| u_t(\tilde{\eta}) - u_t \right\| \right\} + O_p\left(T^{-1/2}\right) \end{aligned} \tag{A.132}$$

where

$$\left\| u_t(\tilde{\eta}) - u_t \right\| \leq \left\| u_t(\tilde{\eta}) - u_t(\eta) \right\| + \left\| u_t(\eta) - u_t \right\|, \tag{A.133}$$

with $\left\| u_t(\eta) - u_t \right\| = O_p\left(\rho^t\right)$. Furthermore, let $\tilde{\Phi}(\bar{p}) = \left[\tilde{\Phi}_0, -\tilde{\Phi}_1, \ldots, -\tilde{\Phi}_{\bar{p}}\right]$, $\Phi(\bar{p}) = \left[\Phi_0, -\Phi_1, \ldots, -\Phi_{\bar{p}}\right]$ and

$Y_t^a(\bar{p}) = \left[y_t^{a\prime} y_{t-1}^{a\prime}, \ldots, y_{t-\bar{p}}^{a\prime}\right]'$. Then

$$\left\|u_t(\tilde{\eta}) - u_t(\eta)\right\| \leq \left\|\sum_{\tau=0}^{t-1}\left[\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta)\right]\tilde{\Phi}(\bar{p})\, Y_{t-\tau}^a(\bar{p})\right\| + \left\|\sum_{\tau=0}^{t-1}\Lambda_\tau(\eta)\left[\tilde{\Phi}(\bar{p}) - \Phi(\bar{p})\right]Y_{t-\tau}^a(\bar{p})\right\|, \quad \text{(A.134)}$$

with

$$\begin{aligned}\left\|\sum_{\tau=0}^{t-1}\left[\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta)\right]\tilde{\Phi}(\bar{p})\, Y_{t-\tau}^a(\bar{p})\right\| &\leq \left\|\tilde{\Phi}(\bar{p})\right\|\left\|Y_t^a(\bar{p})\right\|\left(\sum_{\tau=0}^{t-1}\left\|\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta)\right\|\right) \\ &\leq \frac{C_1}{T^{1/2}}\left(\sum_{\tau=0}^{t-1}\left(1+\kappa^{-1}\right)^{-\tau}\right) = \frac{C_1}{T^{1/2}}\left[\frac{1-\left(1+\kappa^{-1}\right)^{-t}}{1-\left(1+\kappa^{-1}\right)^{-1}}\right] \\ &= O_p\left(T^{-1/2}\right) \qquad \text{(A.135)}\end{aligned}$$

using Lemma **4.1**, and

$$\begin{aligned}\left\|\sum_{\tau=0}^{t-1}\Lambda_\tau(\eta)\left[\tilde{\Phi}(\bar{p}) - \Phi(\bar{p})\right]Y_{t-\tau}^a(\bar{p})\right\| &\leq \left\|\tilde{\Phi}(\bar{p}) - \Phi(\bar{p})\right\|\left\|Y_t^a(\bar{p})\right\|\left(\sum_{\tau=0}^{t-1}\left\|\Lambda_\tau(\eta)\right\|\right) \\ &\leq \frac{C_2}{T^{1/2}}\left(\sum_{\tau=0}^{t-1}\rho^\tau\right)^2 = \frac{C_2}{T^{1/2}}\left(\frac{1-\rho^t}{1-\rho}\right) = O_p\left(T^{-1/2}\right), \qquad \text{(A.136)}\end{aligned}$$

by Theorem **3.3**, for some positive constants C_2 and C_2. Hence

$$\left\|u_t(\tilde{\eta}) - u_t(\eta)\right\| = O_p\left(T^{-1/2}\right) \qquad \text{(A.137)}$$

and then

$$\left\|u_t(\tilde{\eta}) - u_t\right\| = O_p\left(T^{-1/2}\right) + O_p(\rho^t) = O_p\left(T^{-1/2}\right). \qquad \text{(A.138)}$$

Therefore, we get

$$\left\|\tilde{\Sigma}_u(\tilde{\eta}) - \Sigma_u\right\| = O_p\left(T^{-1/2}\right) \qquad \text{(A.139)}$$

and then

$$\left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1}\right\| = O_p\left(T^{-1/2}\right). \qquad \text{(A.140)}$$

□

PROOF OF LEMMA **4.2** Consider the two equations

$$\Phi_0^0 y_t = \mu_{\Phi^0} + \sum_{i=1}^{\bar{p}} \Phi_i^0 y_{t-i} + \Phi_0^0 u_t(\eta^0) + \sum_{j=1}^{\bar{p}} \Theta_j^0 u_{t-j}(\eta^0), \tag{A.141}$$

$$\Phi_0^1 y_t = \mu_{\Phi^1} + \sum_{i=1}^{\bar{p}} \Phi_i^1 y_{t-i} + \Phi_0^1 u_t(\eta^1) + \sum_{j=1}^{\bar{p}} \Theta_j^1 u_{t-j}(\eta^1) \tag{A.142}$$

where $\mu_{\Phi^0} = \Phi^0(1)\mu_y$, $\mu_{\Phi^1} = \Phi^1(1)\mu_y$, with $\Phi^0(1) = \Phi_0^0 - \sum_{i=1}^{\bar{p}} \Phi_i^0$ and $\Phi^1(1) = \Phi_0^1 - \sum_{i=1}^{\bar{p}} \Phi_i^1$. Then substracting (A.140) from (A.141) we get

$$\begin{aligned}\left(\Phi_0^1 - \Phi_0^0\right) v_t(\eta^1) &= (\mu_{\Phi^1} - \mu_{\Phi^0}) + \sum_{i=1}^{\bar{p}} \left(\Phi_i^1 - \Phi_i^0\right) y_{t-i} + \sum_{j=1}^{\bar{p}} \left(\Theta_j^1 - \Theta_j^0\right) u_{t-j}(\eta^1) \\ &\quad + \left(\sum_{j=0}^{\bar{p}} \Theta_j^0 L^j\right)\left[u_t(\eta^1) - u_t(\eta^0)\right] \end{aligned} \tag{A.143}$$

where $v_t(\eta^0) = y_t - u_t(\eta^0)$. Therefore, we have

$$\begin{aligned} u_t(\eta^1) - u_t(\eta^0) &= -\left(\sum_{j=0}^{\bar{p}} \Theta_j^0 L^j\right)^{-1}\left[X_t'(\eta^1) \otimes I_k\right] R(\eta^1 - \eta^0) \\ &= -\sum_{\tau=0}^{t-1}\left[X_{t-\tau}'(\eta^1) \otimes \Lambda_\tau(\eta^0)\right] R(\eta^1 - \eta^0) \end{aligned} \tag{A.144}$$

where $X_t(\eta^1) = \left[1, v_t'(\eta^1), y_{t-1}', \ldots, y_{t-\bar{p}}', u_{t-1}'(\eta^1), \ldots, u_{t-\bar{p}}(\eta^1)\right]'$ with $v_t(\eta^1) = y_t - u_t(\eta^1)$. Finally, we have

$$u_t(\eta^1) - u_t(\eta^0) = -Z_t^\circ(\eta^1, \eta^0)'(\eta^1 - \eta^0) \tag{A.145}$$

where

$$Z_t^\circ(\eta^1, \eta^0) = \sum_{\tau=0}^{t-1} R'\left[X_{t-\tau}(\eta^1) \otimes \Lambda_\tau(\eta^0)'\right]. \tag{A.146}$$

□

PROOF OF PROPOSITION **4.2** Set

$$\bar{Q}_X(\eta) = \left\{\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\Sigma_u^{-1} Z_t(\eta)'\right\}^{-1}. \tag{A.147}$$

Then

$$\begin{aligned}\left\|\tilde{Q}_X^{\circ}(\tilde{\eta})^{-1}-Q_X(\eta)^{-1}\right\|_1 &\leq \left\|\bar{Q}_X^{\circ}(\tilde{\eta})^{-1}-Q_X(\eta)^{-1}\right\| \\ &\leq \left\|\bar{Q}_X^{\circ}(\tilde{\eta})^{-1}-\bar{Q}_X(\eta)^{-1}\right\|+\left\|\bar{Q}_X(\eta)^{-1}-Q_X(\eta)^{-1}\right\|, \qquad \text{(A.148)}\end{aligned}$$

with

$$\begin{aligned}\|\bar{Q}_X(\eta)^{-1}-Q_X(\eta)^{-1}\| &= \left\|\frac{1}{T}\sum_{t=1}^{T}\left\{Z_t(\eta)\,\Sigma_u^{-1}Z_t(\eta)'-\mathsf{E}\left[Z_t(\eta)\,\Sigma_u^{-1}Z_t(\eta)'\right]\right\}\right\| \\ &\leq \|R\|^2\|\Sigma_u^{-1}\|\left\{\sum_{\tau=0}^{\infty}\sum_{v=0}^{\infty}\|\tilde{\Gamma}_X(\tau-v)-\Gamma_X(\tau-\nu)\|\|\Lambda_\tau(\eta)\|\|\Lambda_v(\eta)\|\right\}\end{aligned} \qquad \text{(A.149)}$$

where

$$\tilde{\Gamma}_X(\tau-v)=\frac{1}{T}\sum_{t=1}^{T}X_{t-\tau}X_{t-v}', \quad \Gamma_X(\tau-v)=\mathsf{E}[X_{t-\tau}X_{t-v}']. \qquad \text{(A.150)}$$

From the VARMA structure of y_t one easily sees that

$$\mathsf{E}\|\tilde{\Gamma}_X(\tau-v)-\Gamma_X(\tau-v)\|^2 \leq \frac{\bar{C}\bar{\rho}^{|\tau-v|}}{T}, \qquad \text{(A.151)}$$

for some positive constants $\bar{C}$ and $\rho<\bar{\rho}<1$. Hence

$$\|\bar{Q}_X(\eta)^{-1}-Q_X(\eta)^{-1}\|=O_p\left(T^{-1/2}\right). \qquad \text{(A.152)}$$

Further, it can be easily seen that

$$\left\|\bar{Q}_X^{\circ}(\tilde{\eta})^{-1}-\bar{Q}_X(\eta)^{-1}\right\| \leq \|Q_1\|+\|Q_2\|+\|Q_3\| \qquad \text{(A.153)}$$

where

$$Q_1 = \frac{1}{T}\sum_{t=1}^{T}Z_t(\eta)\,\Sigma_u^{-1}\left[Z_t^{\circ}(\tilde{\eta},\eta)-Z_t(\eta)\right]', \qquad \text{(A.154)}$$

$$Q_2 = \frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1}\right] Z_t^{\circ}(\tilde{\eta},\eta)', \tag{A.155}$$

$$Q_3 = \frac{1}{T}\sum_{t=1}^{T} \left[Z_t^{\circ}(\tilde{\eta},\eta) - Z_t(\eta)\right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} Z_t^{\circ}(\tilde{\eta},\eta)'. \tag{A.156}$$

More especially, we have

$$\|Q_1\| \leq \frac{1}{T}\sum_{t=1}^{T} \|Z_t(\eta)\| \left\|\Sigma_u^{-1}\right\| \|Z_t^{\circ}(\tilde{\eta},\eta) - Z_t(\eta)\|, \tag{A.157}$$

with

$$\begin{aligned}
\mathsf{E}\|Z_t(\eta)\|^2 &= \mathsf{E}\left\|\sum_{\tau=0}^{\infty} R'\left[X_{t-\tau}\otimes \Lambda_\tau(\eta)'\right]\right\|^2 \\
&\leq \|R\|^2 \sum_{\tau_1=0}^{\infty}\sum_{\tau_2=0}^{\infty} \|\Gamma_X(\tau_1-\tau_2)\| \|\Lambda_{\tau_1}(\eta)\| \|\Lambda_{\tau_2}(\eta)\| \\
&\leq \bar{C}_1\|R\|^2 \sum_{\tau_1=0}^{\infty}\sum_{\tau_2=0}^{\infty} \bar{\rho}_1^{|\tau_1-\tau_2|} \|\Lambda_{\tau_1}(\eta)\| \|\Lambda_{\tau_2}(\eta)\| \\
&\leq \bar{C}_2\|R\|^2 \left(\sum_{\tau=0}^{\infty} \|\Lambda_\tau(\eta)\|\right)^2 = O(1),
\end{aligned} \tag{A.158}$$

for some constants $\bar{C}_1, \bar{C}_2 > 0$ and $0 < \rho < \bar{\rho} < 1$, and

$$\begin{aligned}
\|Z_t^{\circ}(\tilde{\eta},\eta) - Z_t(\eta)\| &= \left\|\sum_{\tau=0}^{t-1} R'\left[X_{t-\tau}(\tilde{\eta})\otimes \Lambda_\tau(\eta)'\right] - \sum_{\tau=0}^{\infty} R'\left[X_{t-\tau}\otimes \Lambda_\tau(\eta)'\right]\right\| \\
&\leq \|R\|\left\{\left\|\sum_{\tau=0}^{t-1}\left[\left(X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}\right)\otimes \Lambda_\tau(\eta)'\right]\right\| + \left\|\sum_{\tau=t}^{\infty}\left[X_{t-\tau}\otimes \Lambda_\tau(\eta)'\right]\right\|\right\}
\end{aligned} \tag{A.159}$$

where

$$\begin{aligned}
\mathsf{E}\left\|\sum_{\tau=t}^{\infty}\left[X_{t-\tau}\otimes \Lambda_\tau(\eta)'\right]\right\|^2 &\leq \|R\|^2 \sum_{\tau_1=t}^{\infty}\sum_{\tau_2=t}^{\infty} \|\Gamma_X(\tau_1-\tau_2)\| \|\Lambda_{\tau_1}(\eta)\| \|\Lambda_{\tau_2}(\eta)\| \\
&\leq \bar{C}_1\|R\|^2 \sum_{\tau_1=t}^{\infty}\sum_{\tau_2=t}^{\infty} \bar{\rho}_1^{|\tau_1-\tau_2|} \|\Lambda_{\tau_1}(\eta)\| \|\Lambda_{\tau_2}(\eta)\| \\
&\leq \bar{C}_2\|R\|^2 \left(\sum_{\tau=t}^{\infty} \|\Lambda_\tau(\eta)\|\right)^2 \\
&\leq \bar{C}_2 C\|R\|^2 \left(\sum_{\tau=t}^{\infty} \rho^\tau\right)^2 = O\left(\rho^{2t}\right),
\end{aligned} \tag{A.160}$$

and

$$\left\| \sum_{\tau=0}^{t-1} \left[\left(X_{t-\tau}(\tilde{\eta}) - X_{t-\tau} \right) \otimes \Lambda_\tau(\eta)' \right] \right\| \leq \sum_{\tau=0}^{t-1} \| X_{t-\tau}(\tilde{\eta}) - X_{t-\tau} \| \| \Lambda_\tau(\eta) \|, \tag{A.161}$$

with

$$\| X_{t-\tau}(\tilde{\eta}) - X_{t-\tau} \|^2 = \sum_{j=0}^{\bar{p}} \| u_{t-j-\tau}(\tilde{\eta}) - u_{t-j-\tau} \|^2 = O_p\left(T^{-1}\right) \tag{A.162}$$

in view of (A.137). Therefore

$$\left\| \sum_{\tau=0}^{t-1} \left[\left(X_{t-\tau}(\tilde{\eta}) - X_{t-\tau} \right) \otimes \Lambda_\tau(\eta)' \right] \right\| = O_p\left(T^{-1/2}\right) \tag{A.163}$$

and then

$$\| Z_t^\circ(\tilde{\eta}, \eta) - Z_t(\eta) \| = O_p\left(T^{-1/2}\right) + O_p\left(\rho^t\right) = O_p\left(T^{-1/2}\right). \tag{A.164}$$

Hence

$$\| Q_1 \| = O_p\left(T^{-1/2}\right). \tag{A.165}$$

Likewise, using (A.158), (A.164) and Proposition **4.1**, we can easily show that

$$\begin{aligned} \| Q_2 \| &\leq \left\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \right\| \left\{ \frac{1}{T} \sum_{t=1}^{T} \| Z_t(\eta) \| \left[\| Z_t^\circ(\tilde{\eta}, \eta) - Z_t(\eta) \| + \| Z_t(\eta) \| \right] \right\} \\ &= O_p\left(T^{-1/2}\right), \end{aligned} \tag{A.166}$$

and

$$\begin{aligned} \| Q_3 \| &\leq \frac{1}{T} \sum_{t=1}^{T} \| Z_t^\circ(\tilde{\eta}, \eta) - Z_t(\eta) \| \left\{ \left\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \right\| \| Z_t^\circ(\tilde{\eta}, \eta) - Z_t(\eta) \| \right. \\ &\quad \left. + \| \Sigma_u^{-1} \| \| Z_t^\circ(\tilde{\eta}, \eta) - Z_t(\eta) \| + \left\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \right\| \| Z_t(\eta) \| + \| \Sigma_u^{-1} \| \| Z_t(\eta) \| \right\} \\ &= O_p\left(T^{-1/2}\right). \end{aligned} \tag{A.167}$$

Consequently, we get

$$\left\| \tilde{Q}_X^\circ(\tilde{\eta})^{-1} - \tilde{Q}_X(\eta)^{-1} \right\| = O_p\left(T^{-1/2}\right), \quad \left\| \tilde{Q}_X^\circ(\tilde{\eta})^{-1} - Q_X(\eta)^{-1} \right\| = O_p\left(T^{-1/2}\right), \tag{A.168}$$

and finally

$$\left\|\tilde{Q}_X^{\circ}(\tilde{\eta})^{-1} - Q_X(\eta)^{-1}\right\|_1 = O_p\left(T^{-1/2}\right), \quad \left\|\tilde{Q}_X^{\circ}(\tilde{\eta}) - Q_X(\eta)\right\|_1 = O_p\left(T^{-1/2}\right). \tag{A.169}$$

Furthermore, one can easily show that

$$\begin{aligned}\left\|\tilde{Q}_X(\tilde{\eta})^{-1} - \tilde{Q}_X^{\circ}(\tilde{\eta})^{-1}\right\|_1 &\leq \left\|\tilde{Q}_X(\tilde{\eta})^{-1} - \tilde{Q}_X^{\circ}(\tilde{\eta})^{-1}\right\| \\ &\leq \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\| \frac{1}{T}\sum_{t=1}^{T}\Big\{\left\|Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta},\eta)\right\| \left\|Z_t(\tilde{\eta})\right\| \\ &\quad + \left\|Z_t^{\circ}(\tilde{\eta},\eta)\right\| \left\|Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta},\eta)\right\|\Big\}\end{aligned} \tag{A.170}$$

where, by Proposition **4.1** and Lemma **4.1**,

$$\left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\| \leq \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1}\right\| + \left\|\Sigma_u^{-1}\right\| = O_p(1), \tag{A.171}$$

$$\left\|Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta},\eta)\right\| \leq \|R\| \sum_{\tau=0}^{t-1}\left\{\left\|X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}\right\| + \left\|X_{t-\tau}\right\|\right\} \left\|\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta)\right\| = O_p\left(T^{-1/2}\right), \tag{A.172}$$

and using (A.158), (A.164) and (A.172)

$$\left\|Z_t^{\circ}(\tilde{\eta},\eta)\right\| \leq \left\|Z_t^{\circ}(\tilde{\eta},\eta) - Z_t(\eta)\right\| + \left\|Z_t(\eta)\right\| = O_p(1), \tag{A.173}$$

$$\left\|Z_t(\tilde{\eta})\right\| \leq \left\|Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta},\eta)\right\| + \left\|Z_t^{\circ}(\tilde{\eta},\eta) - Z_t(\eta)\right\| + \left\|Z_t(\eta)\right\| = O_p(1). \tag{A.174}$$

Hence

$$\left\|\tilde{Q}_X(\tilde{\eta})^{-1} - \tilde{Q}_X^{\circ}(\tilde{\eta})^{-1}\right\| = O_p\left(T^{-1/2}\right) \tag{A.175}$$

and then

$$\left\|\tilde{Q}_X(\tilde{\eta})^{-1} - \tilde{Q}_X^{\circ}(\tilde{\eta})^{-1}\right\|_1 = O_p\left(T^{-1/2}\right), \quad \left\|\tilde{Q}_X(\tilde{\eta}) - \tilde{Q}_X^{\circ}(\tilde{\eta})\right\|_1 = O_p\left(T^{-1/2}\right). \tag{A.176}$$

□

PROOF OF THEOREM **4.1** By triangular inequality, we have

$$\begin{aligned}\|\hat{\eta}-\eta\| &\leq \left\|\tilde{Q}_X^{\circ}(\tilde{\eta})\,\tilde{\Omega}_X^{\bullet}(\tilde{\eta})\right\| + \left\|\tilde{Q}_X(\tilde{\eta})\,\tilde{\Omega}_X(\tilde{\eta}) - \tilde{Q}_X^{\circ}(\tilde{\eta})\,\tilde{\Omega}_X^{\circ}(\tilde{\eta})\right\| \\ &\leq \|Q_X(\eta)\|_1\|\Omega_X(\eta)\| + \left\|\tilde{Q}_X^{\circ}(\tilde{\eta}) - Q_X(\eta)\right\|_1\left\|\tilde{\Omega}_X^{\bullet}(\tilde{\eta})\right\| + \|Q_X(\eta)\|_1\left\|\tilde{\Omega}_X^{\bullet}(\tilde{\eta}) - \Omega_X(\eta)\right\| \\ &\quad + \left\|\tilde{Q}_X(\tilde{\eta}) - \tilde{Q}_X^{\circ}(\tilde{\eta})\right\|_1\left\|\tilde{\Omega}_X(\tilde{\eta})\right\| + \left\|\tilde{Q}_X^{\circ}(\tilde{\eta})\right\|_1\left\|\tilde{\Omega}_X(\tilde{\eta}) - \tilde{\Omega}_X^{\circ}(\tilde{\eta})\right\| \qquad \text{(A.177)}\end{aligned}$$

where $\|Q_X(\eta)\|_1 = O_p(1)$, $\|\Omega_X(\eta)\| = O_p\left(T^{-1/2}\right)$,

$$\left\|\tilde{Q}_X^{\circ}(\tilde{\eta}) - Q_X(\eta)\right\|_1 = O_p\left(T^{-1/2}\right), \quad \left\|\tilde{Q}_X(\tilde{\eta}) - \tilde{Q}_X^{\circ}(\tilde{\eta})\right\|_1 = O_p\left(T^{-1/2}\right). \qquad \text{(A.178)}$$

Now, consider $\|\tilde{\Omega}_X^{\bullet}(\tilde{\eta}) - \Omega_X(\eta)\|$ and $\|\tilde{\Omega}_X(\tilde{\eta}) - \tilde{\Omega}_X^{\circ}(\tilde{\eta})\|$. Then, for the first term, we have

$$\begin{aligned}\left\|\tilde{\Omega}_X^{\bullet}(\tilde{\eta}) - \Omega_X(\eta)\right\| &\leq \left\|\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\,\Sigma_u^{-1}\left[u_t(\eta) - u_t\right]\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1}\right]u_t(\eta)\right\| \\ &\quad + \left\|\frac{1}{T}\sum_{t=1}^{T}\left[Z_t^{\circ}(\tilde{\eta},\eta) - Z_t(\eta)\right]\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_t(\eta)\right\| \qquad \text{(A.179)}\end{aligned}$$

where

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\,\Sigma_u^{-1}\left[u_t(\eta) - u_t\right]\right\| &= \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{\infty} R'\left[X_{t-\tau}\otimes\Lambda_\tau(\eta)'\right]\Sigma_u^{-1}\left[u_t(\eta) - u_t\right]\right\| \\ &\leq \|R\|_1\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{\infty}\mathrm{vec}\left[\Lambda_\tau(\eta)'\,\Sigma_u^{-1}\left[u_t(\eta) - u_t\right]X_{t-\tau}'\right]\right\| \\ &= \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{\infty}\Lambda_\tau(\eta)'\,\Sigma_u^{-1}\left[u_t(\eta) - u_t\right]X_{t-\tau}'\right\| \qquad \text{(A.180)}\end{aligned}$$

on using the inequality $\|AB\| \leq \|A\|_1\|B\|$, with $\|R\|_1 = 1$ by construction, and $\|\mathrm{vec}[B]\| = \|B\|$. It follows that

$$\begin{aligned}\mathsf{E}\left\|\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\,\Sigma_u^{-1}\left[u_t(\eta) - u_t\right]\right\| &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{\infty}\|\Lambda_\tau(\eta)\|\|\Sigma_u^{-1}\|\mathsf{E}\left\|\left[u_t(\eta) - u_t\right]X_{t-\tau}'\right\| \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{\infty}\|\Lambda_\tau(\eta)\|\|\Sigma_u^{-1}\|\left\{\mathsf{E}\|u_t(\eta) - u_t\|^2\right\}^{1/2}\left\{\mathsf{E}\|X_{t-\tau}\|^2\right\}^{1/2} \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=0}^{\infty}\|\Lambda_\tau(\eta)\|\right)\|\Sigma_u^{-1}\|\left\{\mathsf{E}\|u_t(\eta) - u_t\|^2\right\}^{1/2}\left\{\mathsf{E}\|X_t\|^2\right\}^{1/2}\end{aligned}$$

$$\leq \bar{C}_3 \|\Sigma_u^{-1}\| \Big(\sum_{\tau=0}^{\infty} \|\Lambda_\tau(\eta)\| \Big) \Big(\frac{1}{T} \sum_{t=1}^{T} \rho^t \Big) = O_p(T^{-1}), \tag{A.181}$$

for some positive constant $\bar{C}_3$. Moreover, we have

$$\begin{aligned} \Big\| \frac{1}{T} \sum_{t=1}^{T} Z_t(\eta) \Big[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \Big] u_t(\eta) \Big\| \leq & \Big\| \frac{1}{T} \sum_{t=1}^{T} Z_t(\eta) \Big[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \Big] [u_t(\eta) - u_t] \Big\| \\ & + \Big\| \frac{1}{T} \sum_{t=1}^{T} Z_t(\eta) \Big[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \Big] u_t \Big\| \end{aligned} \tag{A.182}$$

where similarly as in (A.180) and (A.181), one easily sees that

$$\Big\| \frac{1}{T} \sum_{t=1}^{T} Z_t(\eta) \Big[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \Big] [u_t(\eta) - u_t] \Big\| = O_p(T^{-3/2}). \tag{A.183}$$

Further, manipulating as in (A.180), we can easily show that

$$\begin{aligned} \Big\| \frac{1}{T} \sum_{t=1}^{T} Z_t(\eta) \Big[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \Big] u_t \Big\| &\leq \Big\| \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{\infty} \Lambda_\tau(\eta)' \Big[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \Big] u_t X_{t-\tau}' \Big\| \\ &\leq \Big\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \Big\| \Big\{ \sum_{\tau=0}^{\infty} \|\Lambda_\tau(\eta)\| \Big\| \frac{1}{T} \sum_{t=1}^{T} u_t X_{t-\tau}' \Big\| \Big\} \end{aligned} \tag{A.184}$$

where

$$\Big\{ \sum_{\tau=0}^{\infty} \|\Lambda_\tau(\eta)\| \Big\| \frac{1}{T} \sum_{t=1}^{T} u_t X_{t-\tau}' \Big\| \Big\} = \sum_{\tau=0}^{\infty} \|\Lambda_\tau(\eta)\| \Big\| \frac{1}{T} \sum_{t=1}^{T} u_t X_{t-\tau}' \Big\|, \tag{A.185}$$

and by the VARMA structure of y_t, one easily sees that

$$\Big\| \frac{1}{T} \sum_{t=1}^{T} u_t X_{t-\tau}' \Big\| = O_p\Big(T^{-1/2}\Big). \tag{A.186}$$

Hence

$$\sum_{\tau=0}^{\infty} \|\Lambda_\tau(\eta)\| \Big\| \frac{1}{T} \sum_{t=1}^{T} u_t X_{t-\tau}' \Big\| = O_p\Big(T^{-1/2}\Big). \tag{A.187}$$

Therefore, using Proposition **4.1**, we have

$$\Big\| \frac{1}{T} \sum_{t=1}^{T} Z_t(\eta) \Big[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \Big] u_t \Big\| = O_p\Big(T^{-1}\Big), \tag{A.188}$$

then

$$\left\| \frac{1}{T} \sum_{t=1}^{T} Z_t(\eta) \left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \right] u_t(\eta) \right\| = O_p\left(T^{-3/2}\right) + O_p\left(T^{-1}\right) = O_p\left(T^{-1}\right). \tag{A.189}$$

Finally, one can easily show that

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \left[Z_t^{\circ}(\tilde{\eta}, \eta) - Z_t(\eta) \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t(\eta) \right\| \leq \left\| \Omega_Z^1(\tilde{\eta}) \right\| + \left\| \Omega_Z^2(\tilde{\eta}) \right\| \tag{A.190}$$

where

$$\Omega_Z^1(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \left[Z_t^{\circ}(\tilde{\eta}, \eta) - Z_t(\eta) \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\eta) - u_t \right], \quad \Omega_Z^2(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \left[Z_t^{\circ}(\tilde{\eta}, \eta) - Z_t(\eta) \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t. \tag{A.191}$$

More especially,

$$\left\| \Omega_Z^1(\tilde{\eta}) \right\| \leq \left\| \Omega_Z^{11}(\tilde{\eta}) \right\| + \left\| \Omega_Z^{12}(\tilde{\eta}) \right\| + \left\| \Omega_Z^{13}(\tilde{\eta}) \right\| \tag{A.192}$$

where

$$\Omega_Z^{11}(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} R' \left[X_{t-\tau} \otimes \Lambda_\tau(\eta)' \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\eta) - u_t \right], \tag{A.193}$$

$$\Omega_Z^{12}(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} R' \left[\left\{ X_{t-\tau}(\eta) - X_{t-\tau} \right\} \otimes \Lambda_\tau(\eta)' \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\eta) - u_t \right], \tag{A.194}$$

$$\Omega_Z^{13}(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} R' \left[\left\{ X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}(\eta) \right\} \otimes \Lambda_\tau(\eta)' \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\eta) - u_t \right], \tag{A.195}$$

with $X_t(\eta) = \left[1, v_t'(\eta), y_{t-1}', \ldots, y_{t-\bar{p}}', u_{t-1}'(\eta), \ldots, u_{t-\bar{p}}'(\eta) \right]'$ and $v_t(\eta) = y_t - u_t(\eta)$. And as before we can easily show that

$$\left\| \Omega_Z^{11}(\tilde{\eta}) \right\| \leq \left\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} \right\| \left\{ \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} \left\| \Lambda_\tau(\eta) \right\| \left\| \left[u_t(\eta) - u_t \right] X_{t-\tau}' \right\| \right\} \tag{A.196}$$

where

$$\begin{aligned} \mathsf{E}\left(\frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} \left\| \Lambda_\tau(\eta) \right\| \left\| \left[u_t(\eta) - u_t \right] X_{t-\tau}' \right\| \right) &= \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} \left\| \Lambda_\tau(\eta) \right\| \mathsf{E} \left\| \left[u_t(\eta) - u_t \right] X_{t-\tau}' \right\| \\ &\leq \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} \left\| \Lambda_\tau(\eta) \right\| \left\{ \mathsf{E} \left\| u_t(\eta) - u_t \right\|^2 \right\}^{1/2} \left\{ \mathsf{E} \left\| X_{t-\tau} \right\|^2 \right\}^{1/2} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=t}^{\infty}\|\Lambda_\tau(\eta)\|\Big)\Big\{\mathsf{E}\|u_t(\eta)-u_t\|^2\Big\}^{1/2}\Big\{\mathsf{E}\|X_t\|^2\Big\}^{1/2}\\
&\le \frac{C}{T}\sum_{t=1}^{T}\rho^t\Big(\sum_{\tau=0}^{\infty}\rho^\tau\Big)\Big\{\mathsf{E}\|u_t(\eta)-u_t\|^2\Big\}^{1/2}\Big\{\mathsf{E}\|X_t\|^2\Big\}^{1/2}\\
&\le \frac{\bar{C}_5}{T}\Big(\sum_{t=1}^{T}\rho^{2t}\Big)=\Big[\frac{\bar{C}_5\rho^2\left(1-\rho^{2T}\right)}{T\left(1-\rho^2\right)}\Big]=O\Big(T^{-1}\Big), \qquad \text{(A.197)}
\end{aligned}
$$

for some positive constant $\bar{C}_5$. Hence

$$
\|\Omega_Z^{11}(\tilde{\eta})\| = O_p\Big(T^{-1}\Big). \qquad \text{(A.198)}
$$

Further

$$
\begin{aligned}
\|\Omega_Z^{12}(\tilde{\eta})\| &\le \Big\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\Lambda_\tau(\eta)'\,\tilde{\Sigma}_u(\tilde{\eta})^{-1}\left[u_t(\eta)-u_t\right]\left[X_{t-\tau}(\eta)-X_{t-\tau}\right]'\Big\|\\
&\le \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\Big\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\|\Lambda_\tau(\eta)\|\|u_t(\eta)-u_t\|\|X_{t-\tau}(\eta)-X_{t-\tau}\|\Big\}\\
&\le \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\Big\{\frac{1}{T}\sum_{t=1}^{T}\|u_t(\eta)-u_t\|\Big(\sum_{\tau=0}^{t-1}\|\Lambda_\tau(\eta)\|^2\Big)^{1/2}\Big(\sum_{\tau=0}^{t-1}\|X_{t-\tau}(\eta)-X_{t-\tau}\|^2\Big)^{1/2}\Big\}
\end{aligned}
\qquad \text{(A.199)}
$$

where

$$
\|X_{t-\tau}(\eta)-X_{t-\tau}\|^2=\sum_{j=0}^{\bar{p}}\|u_{t-j-\tau}(\eta)-u_{t-j-\tau}\|^2, \qquad \text{(A.200)}
$$

with

$$
\mathsf{E}\|u_{t-j-\tau}(\eta)-u_{t-j-\tau}\|\le\sum_{v=t-j-\tau}^{\infty}\|\Lambda_v(\eta)\|\|\Phi(\tilde{p})\|\mathsf{E}\|Y_{t-j-\tau-v}^{a}(\tilde{p})\|=O\left(\rho^{t-j-\tau}\right). \qquad \text{(A.201)}
$$

Hence

$$
\|X_{t-\tau}(\eta)-X_{t-\tau}\|^2=O_p\left(\rho^{2(t-\tau)}\right) \qquad \text{(A.202)}
$$

and then

$$
\sum_{\tau=0}^{t-1}\|X_{t-\tau}(\eta)-X_{t-\tau}\|^2=O_p\left(\rho^{2t}\right). \qquad \text{(A.203)}
$$

It follows that

$$
\|\Omega_Z^{12}(\tilde{\eta})\| = O_p\Big(T^{-1}\Big). \qquad \text{(A.204)}
$$

Furthermore, we can easily show that

$$\begin{aligned}\left\|\Omega_Z^{13}(\tilde{\eta})\right\| &\leq \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\Lambda_\tau(\eta)'\,\tilde{\Sigma}_u(\tilde{\eta})^{-1}\left[u_t(\eta)-u_t\right]\left[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right]'\right\| \\ &\leq \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\left\{\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=0}^{t-1}\left\|\Lambda_\tau(\eta)\right\|\left\|u_t(\eta)-u_t\right\|\left\|X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right\|\right)\right\} \quad \text{(A.205)}\end{aligned}$$

where

$$\left\|X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right\|^2=\sum_{j=0}^{\bar{p}}\left\|u_{t-j-\tau}(\tilde{\eta})-u_{t-j-\tau}(\eta)\right\|^2, \quad \text{(A.206)}$$

with

$$\left\|u_{t-j-\tau}(\tilde{\eta})-u_{t-j-\tau}(\eta)\right\|=\left\|u_t(\tilde{\eta})-u_t(\eta)\right\|=O_p\left(T^{-1/2}\right), \quad \text{(A.207)}$$

using (A.137). As a result, we get

$$\left\|\Omega_Z^{13}(\tilde{\eta})\right\|=O_p\left(T^{-3/2}\right) \quad \text{(A.208)}$$

since $\left\|u_t(\eta)-u_t\right\|_i=O_p\left(\rho^t\right)$, and consequently

$$\left\|\Omega_Z^{1}(\tilde{\eta})\right\|=\left\|\frac{1}{T}\sum_{t=1}^{T}\left[Z_t^{\circ}(\tilde{\eta},\eta)-Z_t(\eta)\right]\tilde{\Sigma}_u(\tilde{\eta})^{-1}\left[u_t(\eta)-u_t\right]\right\|=O_p\left(T^{-1}\right) \quad \text{(A.209)}$$

In addition, one easily sees that

$$\left\|\Omega_Z^{2}(\tilde{\eta})\right\|\leq\left\|\Omega_Z^{21}(\tilde{\eta})\right\|+\left\|\Omega_Z^{22}(\tilde{\eta})\right\|+\left\|\Omega_Z^{23}(\tilde{\eta})\right\| \quad \text{(A.210)}$$

where

$$\Omega_Z^{21}(\tilde{\eta}) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}R'\left[X_{t-\tau}\otimes\Lambda_\tau(\eta)'\right]\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_t, \quad \text{(A.211)}$$

$$\Omega_Z^{22}(\tilde{\eta}) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}R'\left[\left\{X_{t-\tau}(\eta)-X_{t-\tau}\right\}\otimes\Lambda_\tau(\eta)'\right]\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_t, \quad \text{(A.212)}$$

$$\Omega_Z^{23}(\tilde{\eta}) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}R'\left[\left\{X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right\}\otimes\Lambda_\tau(\eta)'\right]\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_t, \quad \text{(A.213)}$$

Likewise, one can easily show that

$$\|\Omega_Z^{21}(\tilde{\eta})\| \leq \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\| \left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}\|\Lambda_\tau(\eta)\|\|u_t X_{t-\tau}'\|\right\} \tag{A.214}$$

where by independence between u_t and X_t

$$\begin{aligned}
\mathsf{E}\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}\|\Lambda_\tau(\eta)\|\|u_t X_{t-\tau}'\|\right\} &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}\|\Lambda_\tau(\eta)\|\left\{\mathsf{E}\|u_t X_{t-\tau}'\|^2\right\}^{1/2} \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}\|\Lambda_\tau(\eta)\|\left\{\mathsf{E}\|u_t\|^2\mathsf{E}\|X_{t-\tau}\|^2\right\}^{1/2} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=t}^{\infty}\|\Lambda_\tau(\eta)\|\right)\left\{\mathsf{E}\|u_t\|^2\mathsf{E}\|X_t\|^2\right\}^{1/2} \\
&= O\left(T^{-1}\right),
\end{aligned} \tag{A.215}$$

then

$$\|\Omega_Z^{21}(\tilde{\eta})\| = O_p\left(T^{-1}\right). \tag{A.216}$$

In addition, as for (A.194) we can easily show using (A.203), that

$$\begin{aligned}
\|\Omega_Z^{22}(\tilde{\eta})\| &\leq \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\Lambda_\tau(\eta)'\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_t\left[X_{t-\tau}(\eta)-X_{t-\tau}\right]'\right\| \\
&\leq \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\|\Lambda_\tau(\eta)\|\|u_t\|\|X_{t-\tau}(\eta)-X_{t-\tau}\|\right\} \\
&\leq \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\left\{\frac{1}{T}\sum_{t=1}^{T}\|u_t\|\left(\sum_{\tau=0}^{t-1}\|\Lambda_\tau(\eta)\|^2\right)^{1/2}\left(\sum_{\tau=0}^{t-1}\|X_{t-\tau}(\eta)-X_{t-\tau}\|^2\right)^{1/2}\right\} \\
&= O_p\left(T^{-1}\right).
\end{aligned} \tag{A.217}$$

Finally

$$\begin{aligned}
\|\Omega_Z^{23}(\tilde{\eta})\| &\leq \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\Lambda_\tau(\eta)'\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_t\left[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right]'\right\| \\
&= \left\|\frac{1}{T}\sum_{\tau=0}^{T-1}\sum_{t=\tau+1}^{T}\Lambda_\tau(\eta)'\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_t\left[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right]'\right\| \\
&\leq \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\left\{\sum_{\tau=0}^{T-1}\|\Lambda_\tau(\eta)\|\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_t\left[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right]'\right\|\right\}
\end{aligned} \tag{A.218}$$

where

$$\left\| \frac{1}{T} \sum_{t=\tau+1}^{T} u_t \left[X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}(\eta) \right]' \right\|^2 = \sum_{j=0}^{\bar{p}} \left\| \frac{1}{T} \sum_{t=\tau+1}^{T} u_t \left[u_{t-j-\tau}(\tilde{\eta}) - u_{t-j-\tau}(\eta) \right]' \right\|^2, \tag{A.219}$$

with

$$\begin{aligned} \left\| \frac{1}{T} \sum_{t=\tau+1}^{T} u_t \left[u_{t-\tau}(\tilde{\eta}) - u_{t-\tau}(\eta) \right]' \right\| &= \left\| \frac{1}{T} \sum_{t=\tau+1}^{T} \sum_{v=0}^{t-\tau-1} u_t Y_{t-\tau-v}^{a\prime} \left[\tilde{\Phi}(\bar{p})' \Lambda_v(\tilde{\eta}) - \Phi(\bar{p})' \Lambda_v(\eta) \right] \right\| \\ &= \left\| \frac{1}{T} \sum_{v=0}^{T-\tau-1} \sum_{t=\tau+1+v}^{T} u_t Y_{t-\tau-v}^{a\prime} \left[\tilde{\Phi}(\bar{p})' \Lambda_v(\tilde{\eta}) - \Phi(\bar{p})' \Lambda_v(\eta) \right] \right\| \\ &\leq \sum_{v=0}^{T-\tau-1} \left\| \frac{1}{T} \sum_{t=\tau+1+v}^{T} u_t Y_{t-\tau-v}^{a\prime} \right\| \left\| \tilde{\Phi}(\bar{p})' \Lambda_v(\tilde{\eta}) - \Phi(\bar{p})' \Lambda_v(\eta) \right\| \end{aligned} \tag{A.220}$$

where by independence between u_t and Y_t^a

$$\left\| \frac{1}{T} \sum_{t=\tau+1+v}^{T} u_t Y_{t-\tau-v}^{a\prime} \right\| = O_p\left(T^{-1/2}\right). \tag{A.221}$$

Moreover, using Theorem (**3.3**) and Lemma (**4.1**), we have

$$\begin{aligned} \sum_{v=0}^{T-\tau-1} \left\| \tilde{\Phi}(\bar{p})' \Lambda_v(\tilde{\eta}) - \Phi(\bar{p})' \Lambda_v(\tilde{\eta}) \right\| &\leq \sum_{v=0}^{T-\tau-1} \left\| \tilde{\Phi}(\bar{p}) \right\| \left\| \Lambda_v(\tilde{\eta}) - \Lambda_v(\eta) \right\| + \left\| \tilde{\Phi}(\bar{p}) - \Phi(\bar{p}) \right\| \left\| \Lambda_v(\eta) \right\| \\ &= O_p\left(T^{-1/2}\right). \end{aligned} \tag{A.222}$$

Then

$$\left\| \frac{1}{T} \sum_{t=\tau+1}^{T} u_t \left[u_{t-\tau}(\tilde{\eta}) - u_{t-\tau}(\eta) \right]' \right\| = \left\| \frac{1}{T} \sum_{t=\tau+1}^{T} u_t \left[u_{t-j-\tau}(\tilde{\eta}) - u_{t-j-\tau}(\eta) \right]' \right\| = O_p\left(T^{-1}\right), \tag{A.223}$$

$$\left\| \frac{1}{T} \sum_{t=\tau+1}^{T} u_t \left[X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}(\eta) \right]' \right\| = O_p\left(T^{-1}\right), \tag{A.224}$$

and consequently

$$\left\| \Omega_Z^{23}(\tilde{\eta}) \right\| = O_p\left(T^{-1}\right). \tag{A.225}$$

Therefore, we get

$$\|\Omega_Z^2(\tilde{\eta})\| = \left\| \frac{1}{T} \sum_{t=1}^{T} \left[Z_t^{\circ}(\tilde{\eta}, \eta) - Z_t(\eta) \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t \right\| = O_p\left(T^{-1}\right), \tag{A.226}$$

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \left[Z_t^{\circ}(\tilde{\eta}, \eta) - Z_t(\eta) \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t(\eta) \right\| = O_p\left(T^{-1}\right) \tag{A.227}$$

and finally

$$\left\| \tilde{\Omega}_X^{\bullet}(\tilde{\eta}) - \Omega_X(\eta) \right\| = O_p(T^{-1}). \tag{A.228}$$

Similarly, one can easily see that

$$\left\| \tilde{\Omega}_X(\tilde{\eta}) - \tilde{\Omega}_X^{\circ}(\tilde{\eta}) \right\| \leq \|R\| \left\{ \|\Omega_X^1(\tilde{\eta})\| + \|\Omega_X^2(\tilde{\eta})\| + \|\Omega_X^3(\tilde{\eta})\| + \|\Omega_X^4(\tilde{\eta})\| \right\} \tag{A.229}$$

where

$$\Omega_X^1(\tilde{\eta}) = R' \mathrm{vec} \left[\frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} \left[\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta) \right]' \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\tilde{\eta}) - u_t \right] \left[X_{t-\tau}(\tilde{\eta}) - X_{t-\tau} \right]' \right], \tag{A.230}$$

$$\Omega_X^2(\tilde{\eta}) = R' \mathrm{vec} \left[\frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} \left[\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta) \right]' \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\tilde{\eta}) - u_t \right] X_{t-\tau}' \right], \tag{A.231}$$

$$\Omega_X^3(\tilde{\eta}) = R' \mathrm{vec} \left[\frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} \left[\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta) \right]' \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t \left[X_{t-\tau}(\tilde{\eta}) - X_{t-\tau} \right]' \right], \tag{A.232}$$

$$\Omega_X^4(\tilde{\eta}) = R' \mathrm{vec} \left[\frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} \left[\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta) \right]' \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t X_{t-\tau}' \right]. \tag{A.233}$$

Using the same arguments as before, one easily sees that

$$\begin{aligned}
\|\Omega_X^1(\tilde{\eta})\| &\leq \left\| \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} \left[\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta) \right]' \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\tilde{\eta}) - u_t \right] \left[X_{t-\tau}(\tilde{\eta}) - X_{t-\tau} \right]' \right\| \\
&\leq \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} \left\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} \right\| \|\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta)\| \Big\{ \|u_t(\tilde{\eta}) - u_t(\eta)\| \|X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}(\eta)\| \\
&\quad + \|u_t(\eta) - u_t\| \|X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}(\eta)\| + \|u_t(\tilde{\eta}) - u_t(\eta)\| \|X_{t-\tau}(\eta) - X_{t-\tau}\| \\
&\quad + \|u_t(\eta) - u_t\| \|X_{t-\tau}(\eta) - X_{t-\tau}\| \Big\} \\
&= O_p\left(T^{-3/2}\right) + O_p\left(T^{-2}\right) + O_p\left(T^{-2}\right) + O_p\left(T^{-3/2}\right) = O_p\left(T^{-3/2}\right),
\end{aligned} \tag{A.234}$$

$$\begin{aligned}
\|\Omega_X^2(\tilde{\eta})\| &\leq \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\left[\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\dot{\eta})\right]'\tilde{\Sigma}_u(\tilde{\eta})^{-1}\left[u_t(\tilde{\eta})-u_t\right]X_{t-\tau}'\right\| \\
&\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\|\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\|\Big\{\|u_t(\tilde{\eta})-u_t(\eta)\|\|X_{t-\tau}\| \\
&\quad +\|u_t(\eta)-u_t\|\|X_{t-\tau}\|\Big\} \\
&= O_p(T^{-1})+O_p(T^{-3/2})=O_p(T^{-1}), \qquad \text{(A.235)}
\end{aligned}$$

$$\begin{aligned}
\|\Omega_X^3(\tilde{\eta})\| &\leq \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\left[\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\right]'\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_t\left[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\right]'\right\| \\
&\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\|\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\|\Big\{\|u_t\|\|X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\| \\
&\quad +\|u_t\|\|X_{t-\tau}(\eta)-X_{t-\tau}\|\Big\} \\
&= O_p(T^{-1})+O_p(T^{-3/2})=O_p(T^{-1}) \qquad \text{(A.236)}
\end{aligned}$$

and

$$\begin{aligned}
\|\Omega_X^4(\tilde{\eta})\| &\leq \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\left[\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\right]'\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_tX_{t-\tau}'\right\| \\
&\leq \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\left\{\sum_{\tau=0}^{T-1}\|\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\|\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_tX_{t-\tau}'\right\|\right\}=O_p(T^{-1}). \qquad \text{(A.237)}
\end{aligned}$$

Hence

$$\left\|\tilde{\Omega}_X(\tilde{\eta})-\tilde{\Omega}_X^{\circ}(\tilde{\eta})\right\|=O_p(T^{-1}). \qquad \text{(A.238)}$$

In particular, one can easily see that

$$\begin{aligned}
\left\|\tilde{\Omega}_X(\tilde{\eta})\right\| &\leq \left\|\tilde{\Omega}_X^{\circ}(\tilde{\eta})\right\|+\left\|\tilde{\Omega}_X(\tilde{\eta})-\tilde{\Omega}_X^{\circ}(\tilde{\eta})\right\| \\
&\leq \left\|\Omega_X(\eta)\right\|+\left\|\tilde{\Omega}_X^{\circ}(\tilde{\eta})-\Omega_X(\eta)\right\|+\left\|\tilde{\Omega}_X(\tilde{\eta})-\tilde{\Omega}_X^{\circ}(\tilde{\eta})\right\| \qquad \text{(A.239)}
\end{aligned}$$

where, using the same arguments as before it can be easily shown that

$$\left\|\tilde{\Omega}_X^{\circ}(\tilde{\eta})-\Omega_X(\eta)\right\|=O_p(T^{-1}), \qquad \text{(A.240)}$$

and then

$$\left\| \tilde{\Omega}_X(\tilde{\eta}) \right\| = O_p(T^{-1/2}) \tag{A.241}$$

Hence we conclude that

$$\|\hat{\eta} - \eta\| = O_p(T^{-1/2}). \tag{A.242}$$

□

PROOF OF PROPOSITION **4.3** Recall that

$$\begin{aligned} \left\| \tilde{S}_X(\tilde{\eta}) - S_X(\eta) \right\| \leq\ & T^{1/2} \Big\{ \left\| \tilde{Q}_X^{\circ}(\tilde{\eta}) - Q_X(\eta) \right\|_1 \left\| \tilde{\Omega}_X^{\bullet}(\tilde{\eta}) \right\| + \|Q_X(\eta)\|_1 \left\| \tilde{\Omega}_X^{\bullet}(\tilde{\eta}) - \Omega_X(\eta) \right\| \\ & + \left\| \tilde{Q}_X(\tilde{\eta}) - \tilde{Q}_X^{\circ}(\tilde{\eta}) \right\|_1 \left\| \tilde{\Omega}_X(\tilde{\eta}) \right\| + \left\| \tilde{Q}_X^{\circ}(\tilde{\eta}) \right\|_1 \left\| \tilde{\Omega}_X(\tilde{\eta}) - \tilde{\Omega}_X^{\circ}(\tilde{\eta}) \right\| \Big\}. \end{aligned} \tag{A.243}$$

Then it follows by Proposition **4.2** and Theorem **4.1** that

$$\left\| \tilde{S}_X(\tilde{\eta}) - S_X(\eta) \right\| = O_p(T^{-1/2}). \tag{A.244}$$

□

PROOF OF THEOREM **4.2** By the central limit theorem for stationary processes [see Anderson (1971, Section 7.7), Scott (1973, Theorem 2) and Chung (2001, Theorem 9.1.5)] and under the assumption of independence between u_t and $Z_t(\eta)$, we have

$$T^{1/2} \Omega_X(\eta) \underset{T \to \infty}{\longrightarrow} N\left[0, Q_X(\eta)^{-1}\right]. \tag{A.245}$$

Then by Proposition **4.3**, we get

$$T^{1/2}(\hat{\eta} - \eta) = \tilde{S}_X(\tilde{\eta}) \underset{T \to \infty}{\longrightarrow} N\left[0, Q_X(\eta)\right]. \tag{A.246}$$

□

PROOF OF PROPOSITION **5.1** Let n_{T_1} and $P^1 = (p_1^1, \ldots, p_k^1)'$ be the estimates of n_{T_0} and $P^0 = (p_1^0, \ldots, p_k^0)'$ that minimize information criteria (5.10) and (5.11), respectively. Let, also, $\bar{p}^1 = \max(p_1^1, \ldots, p_k^1)$ and $\bar{p}^0 = \max(p_1^0, \ldots, p_k^0)$, then the number of the echelon form freely varying parameters corresponding to P^1 and P^0 are $r_{\bar{p}^1}$ and $r_{\bar{p}^0}$, respectively. For the first stage, we have

$$\ln\Big(\det\{\tilde{\Sigma}_u(n_{T_0})\}\Big) - \ln\Big(\det\{\tilde{\Sigma}_u(n_{T_1})\}\Big) + c_1 k^2 \frac{(n_{T_0} - n_{T_1})}{T^{1/2}} + c_1^2 k^4 \frac{(n_{T_0}^2 - n_{T_1}^2)}{T} \leq 0, \tag{A.247}$$

Assume now that $n_{T_1} < n_{T_0}$. It follows that

$$\ln\Big(\det\{\tilde{\Sigma}_u(n_{T_0})\}\Big) - \ln\Big(\det\{\tilde{\Sigma}_u(n_{T_1})\}\Big) < 0, \tag{A.248}$$

and as T goes to infinity $(n_{T_0} - n_{T_1})/T^{1/2}$ and $(n_{T_0}^2 - n_{T_1}^2)/T$ vanish and the minimum of the information criterion will not be reached. So the probability to choose too small an order goes to zero as $T \to \infty$ and n_{T_1} must equal n_{T_0}. Inversely, if $n_{T_1} > n_{T_0}$ then

$$\ln\Big(\det\{\tilde{\Sigma}_u(n_{T_0})\}\Big) - \ln\Big(\det\{\tilde{\Sigma}_u(n_{T_1})\}\Big) > 0, \tag{A.249}$$

and as $T \to \infty$ the penalty term of the lower order will be smaller than the greater one. So the probability of choosing the small order approaches one in large samples. Hence as T goes to infinity $n_{T_1} \to n_{T_0}$. Similarly, consider

$$\ln\Big(\det\{\tilde{\Sigma}_e(P^0)\}\Big) - \ln\Big(\det\{\tilde{\Sigma}_e(P^1)\}\Big) + c_2(r_{\bar{p}^0} - r_{\bar{p}^1})\frac{\hat{n}_T}{T^{1/2}} + c_2^2(r_{\bar{p}^0}^2 - r_{\bar{p}^1}^2)\frac{\hat{n}_T^2}{T} \leq 0, \tag{A.250}$$

for the second stage, where for example $\tilde{\Sigma}_e(P^1) = \tilde{\Sigma}_e(\hat{n}_T)$ such that the vector of Kronecker indices considered in the second stage is $P = P^1$. Now, suppose that $P^1 < P^0$ then $r_{\bar{p}^1} < r_{\bar{p}^0}$ and

$$\ln\Big(\det\{\tilde{\Sigma}_e(P^0)\}\Big) - \ln\Big(\det\{\tilde{\Sigma}_e(P^1)\}\Big) < 0. \tag{A.251}$$

Again, the same argument applies and as T goes to infinity $\hat{n}_T/T^{1/2} \to 0$ then in large samples the probability

of choosing too small an order falls to zero and P^1 must equal to P^0 to reach the minimum. While in the case where $P^1 > P^0$ and $r_{\tilde{p}^1} > r_{\tilde{p}^0}$ we have

$$\ln\Big(\det\big\{\tilde{\Sigma}_e\left(P^0\right)\big\}\Big) - \ln\Big(\det\big\{\tilde{\Sigma}_e\left(P^1\right)\big\}\Big) > 0, \tag{A.252}$$

so that in large samples

$$\frac{\ln\Big(\det\big\{\tilde{\Sigma}_e\left(P^0\right)\big\}\Big)}{\ln\Big(\det\big\{\tilde{\Sigma}_e\left(P^1\right)\big\}\Big)} \longrightarrow 1. \tag{A.253}$$

Moreover, the penalty term associated with $r_{\tilde{p}^0}$ falls more quickly to zero than that corresponding to $r_{\tilde{p}^1}$ and small order of the Kronecker indices are more likely to be chosen. Hence, as T goes to infinity P^1 converges to P^0. □

B. Appendix : Chosen value of the Cholesky matrix P_u

The chosen values for the lower triangular matrix P_u in the simulation study, were

$$P_u = \begin{bmatrix} 0.7 & 0.0 \\ -0.2 & 0.5 \end{bmatrix}, \text{ for } k = 2,$$

$$P_u = \begin{bmatrix} 0.7 & 0.0 & 0.0 \\ -0.2 & 0.5 & 0.0 \\ 0.4 & -0.7 & 0.8 \end{bmatrix}, \text{ for } k = 3,$$

$$P_u = \begin{bmatrix} 0.7 & 0.0 & 0.0 & 0.0 \\ -0.2 & 0.5 & 0.0 & 0.0 \\ 0.4 & -0.7 & 0.8 & 0.0 \\ -0.8 & 0.9 & 0.2 & 0.9 \end{bmatrix}, \text{ for } k = 4,$$

$$P_u = \begin{bmatrix} 0.7 & 0.0 & 0.0 & 0.0 & 0.0 \\ -0.2 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.4 & -0.7 & 0.8 & 0.0 & 0.0 \\ -0.8 & 0.9 & 0.2 & 0.9 & 0.0 \\ -0.9 & -0.6 & -0.8 & 0.5 & 0.7 \end{bmatrix}, \text{ for } k = 5.$$

C. Appendix : The estimated echelon VARMA model

The estimated model is an echelon VARMA model with Kronecker indices $(5,4,2,2,3,4)$ such that

$$y_t = \tilde{\mu}_{\Phi} + \left(I_k - \tilde{\Phi}_0\right)\tilde{v}_t\left(n_T\right) + \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i y_{t-i} + \sum_{j=1}^{\bar{p}} \tilde{\Theta}_j \tilde{u}_{t-j}\left(n_T\right) + \tilde{u}_t\left(n_T\right)$$

where $n_T = 12$, $\tilde{v}_t\left(n_T\right) = y_t - \tilde{u}_t\left(n_T\right)$, $\bar{p} = \max\left(5,4,2,2,3,4\right) = 5$,

$$\tilde{\mu}_{\Phi} = \begin{bmatrix} 0.0026 \\ 0.0078 \\ -0.1104 \\ -0.0006 \\ -0.0038 \\ 0.0076 \end{bmatrix}, \quad \tilde{\Phi}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1.3273 & 1 & 0 & 0 & 0 & 0 \\ -18.1708 & -0.1760 & 1 & 0 & 0 & 0 \\ 0.4064 & 0.0279 & 0 & 1 & 0 & 0 \\ -0.1190 & 0.0292 & 0 & 0 & 1 & 0 \\ 1.2814 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\tilde{\Phi}_1 = \begin{bmatrix} -0.0216 & 0 & 0 & 0 & 0 & 0 \\ 0.0325 & 2.5877 & 0 & 0 & 0 & -2.1977 \\ -13.4978 & -2.6238 & 0.5795 & -2.0254 & 4.8679 & 2.8881 \\ 0.2996 & 0.6697 & -0.0089 & -0.2119 & 0.4815 & -0.6438 \\ -0.0089 & -0.9558 & 0 & 0 & -0.1240 & 1.0588 \\ 0.0777 & 2.5104 & 0 & 0 & 0 & -2.1130 \end{bmatrix},$$

$$\tilde{\Phi}_2 = \begin{bmatrix} -0.0025 & 0.2313 & 0 & 0 & 0 & -0.1725 \\ -0.4746 & 0.5206 & 0 & 0 & 0.1526 & -0.4283 \\ 18.4278 & -8.4691 & -0.3796 & -3.3857 & 3.4627 & 15.2337 \\ 0.0553 & -0.9047 & 0.0062 & 0.1593 & 0.0656 & 0.7850 \\ 0.3861 & -0.4065 & -0.0084 & -0.2344 & 0.3500 & 0.3546 \\ -0.4924 & 0.3819 & 0 & 0 & 0.1090 & -0.3076 \end{bmatrix},$$

$$\tilde{\Phi}_3 = \begin{bmatrix} 0.0784 & -0.5348 & 0 & 0 & 0.0178 & 0.5032 \\ -0.0151 & -0.6390 & 0.0035 & 0.0418 & -0.0939 & 0.4406 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0915 & 0.8070 & -0.0005 & 0.0375 & 0.4388 & -0.8210 \\ 0.0175 & -1.3704 & 0.0042 & 0.0460 & -0.0881 & 1.2022 \end{bmatrix},$$

$$\tilde{\Phi}_4 = \begin{bmatrix} 0.0384 & 0.2752 & -0.0025 & -0.1046 & 0.2355 & -0.2238 \\ 0.3908 & -0.5459 & 0.0004 & 0.0075 & 0.3055 & 0.9008 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3985 & -0.5566 & 0.0006 & 0.0287 & 0.2887 & 0.8909 \end{bmatrix},$$

$$\tilde{\Phi}_5 = \begin{bmatrix} 0.0584 & 0.2261 & -0.0002 & -0.1333 & 0.0295 & -0.2758 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\tilde{\Theta}_1 = \begin{bmatrix} -0.3621 & 0.0497 & -0.0003 & -0.0592 & 0.0010 & -0.0480 \\ 0.1163 & -2.9859 & -0.0010 & -0.0929 & 0.0923 & 3.0130 \\ 20.0320 & -9.8960 & -0.3735 & -10.0991 & 4.4153 & 15.8095 \\ -0.4179 & -0.8620 & 0.0043 & 0.0400 & -0.1554 & 0.7377 \\ 0.1722 & 0.9292 & -0.0020 & -0.0575 & 0.3331 & -1.0415 \\ 0.0890 & -2.8739 & -0.0007 & -0.0871 & 0.1062 & 2.9033 \end{bmatrix},$$

$$\tilde{\Theta}_2 = \begin{bmatrix} 0.0083 & 0.1193 & -0.0006 & 0.0338 & 0.0820 & -0.1558 \\ 0.6163 & 0.8915 & -0.0075 & 0.0503 & 0.0003 & -0.9225 \\ -6.6873 & 21.5106 & 0.2097 & 8.0199 & -6.8990 & -26.6069 \\ -0.1960 & 0.7956 & -0.0051 & -0.4257 & 0.2565 & -0.7354 \\ -0.2730 & 0.7494 & 0.0057 & 0.1822 & -0.3240 & -0.7977 \\ 0.6386 & 0.9455 & -0.0074 & 0.0526 & 0.0455 & -0.9583 \end{bmatrix},$$

$$\tilde{\Theta}_3 = \begin{bmatrix} 0.0444 & 0.2686 & 0.0007 & 0.0284 & -0.0884 & -0.2416 \\ 0.2975 & 0.2605 & -0.0021 & -0.0636 & -0.2176 & -0.1928 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -0.1013 & -0.6037 & 0.0006 & -0.2053 & -0.1459 & 0.5602 \\ 0.2342 & 0.8055 & -0.0029 & -0.0598 & -0.2130 & -0.7476 \end{bmatrix},$$

$$\tilde{\Theta}_4 = \begin{bmatrix} -0.0354 & -0.6877 & 0.0030 & 0.2105 & -0.2812 & 0.6214 \\ -0.3901 & -0.2943 & -0.0033 & 0.1991 & -0.1798 & -0.0197 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -0.4104 & -0.2102 & -0.0043 & 0.1504 & -0.1204 & -0.0994 \end{bmatrix},$$

$$\tilde{\Theta}_5 = \begin{bmatrix} -0.0638 & 0.1785 & -0.0004 & 0.1062 & -0.0860 & -0.0819 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

with

$$\tilde{\mu}_y = \begin{bmatrix} 0.0022 \\ -0.0012 \\ -0.0006 \\ -0.0003 \\ -0.0005 \\ 0.0025 \end{bmatrix},$$

and

$$P_e = \begin{bmatrix} 0.00604 & 0 & 0 & 0 & 0 & 0 \\ 0.00033 & 0.01804 & 0 & 0 & 0 & 0 \\ 0.06575 & 0.05219 & 0.54088 & 0 & 0 & 0 \\ -0.00051 & -0.00262 & -0.00423 & 0.01345 & 0 & 0 \\ -0.00005 & -0.00067 & 0.00172 & 0.00593 & 0.00706 & 0 \\ 0.00035 & 0.01796 & 0.00005 & 0.00023 & 0.00011 & 0.00180 \end{bmatrix}$$

where $\tilde{\mu}_y = \left[\tilde{\Phi}(1)\right]^{-1} \tilde{\mu}_{\Phi}$, $\tilde{\Phi}(1) = \tilde{\Phi}_0 - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i$, and such that $P_e P_e' = \tilde{\Sigma}_e(n_T)$. Note here, that the McMillan degree is $\sum_{l=1}^{k} p_l = 20$, and the matrix formed by the estimated Kronecker indices is

$$\bar{P} = \begin{bmatrix} 5 & 4 & 2 & 2 & 3 & 4 \\ 5 & 4 & 2 & 2 & 3 & 4 \\ 3 & 3 & 2 & 2 & 2 & 2 \\ 3 & 3 & 2 & 2 & 2 & 2 \\ 4 & 4 & 2 & 2 & 3 & 3 \\ 5 & 4 & 2 & 2 & 3 & 4 \end{bmatrix}.$$

This leads to $\sum_{l=1}^{k} \sum_{m=1}^{k} p_{lm} = 106$ autoregressive and $k \sum_{l=1}^{k} p_l = 120$ moving average free coefficients, respectively. Hence, 232 overall model free parameters, including the 6×1 vector of constants. Which is almost about half of 438 the number of parameters implied by a sixvariate VAR(12) with constant terms.

Table xi. Frequencies of estimated Kronecker indices for several bivariate ARMA models with different sample sizes

(p_1, p_2)	$\lambda_{\max}^{AR}$	$\lambda_{\max}^{MA}$	T=100	T=200	T=300	T=400	T=500	T=1000	T=2000
$(0,1)^*$	0.9000	0.7000	957	989	988	997	1000	999	1000
$(1,1)$	-	-	2	0	0	0	0	0	0
$(0,2)$	-	-	38	11	12	3	0	1	0
$(0,3)$	-	-	3	0	0	0	0	0	0
$(1,1)^*$	0.9000	0.5949*	997	998	998	1000	1000	1000	1000
$(1,2)$	-	-	1	2	0	0	0	0	0
$(2,1)$	-	-	2	0	2	0	0	0	0
$(1,0)^*$	0.9000	0.7000	980	996	995	999	1000	1000	1000
$(1,1)$	-	-	4	0	0	0	0	0	0
$(2,0)$	-	-	16	4	5	1	0	0	0
$(0,2)^*$	0.9848	0.9219*	813	869	871	907	912	938	976
(0.3)	-	-	172	124	116	81	69	36	7
$(1,2)$	-	-	9	4	12	11	15	24	17
$(0,4)$	-	-	4	1	0	0	0	1	0
$(1,3)$	-	-	2	2	1	1	4	1	0
$(1,2)^*$	0.9000	0.8240	990	996	1000	999	1000	1000	1000
$(1,1)$	-	-	3	0	0	0	0	0	0
$(1,3)$	-	-	2	2	0	1	0	0	0
$(2,2)$	-	-	5	2	0	0	0	0	0
$(2,2)^*$	0.7275	0.7891*	539	752	915	956	980	999	1000
$(2,1)$	-	-	438	233	77	34	18	0	0
$(1,2)$	-	-	5	0	0	0	0	0	0
$(2,3)$	-	-	5	9	7	6	2	1	0
$(3,2)$	-	-	13	6	1	4	0	0	0
$(2,1)^*$	0.9000	0.6806*	992	993	995	1000	1000	1000	1000
$(2,0)$	-	-	1	0	0	0	0	0	0
$(2,2)$	-	-	1	2	2	0	0	0	0
$(3,0)$	-	-	1	0	0	0	0	0	0
$(3,1)$	-	-	5	5	3	0	0	0	0
$(2,0)^*$	0.9848*	0.4000	949	978	980	990	998	1000	1000
$(2,1)$	-	-	44	17	16	9	2	0	0
$(2,2)$	-	-	2	0	0	0	0	0	0
$(3,0)$	-	-	5	5	4	1	0	0	0
$(0,3)^*$	0.9848*	0.9547*	944	982	984	981	990	982	979
$(0,4)$	-	-	42	9	3	2	0	1	0
$(1,3)$	-	-	12	9	13	16	10	17	19
$(1,4)$	-	-	1	0	0	0	0	0	2
$(2,3)$	-	-	1	0	0	1	0	0	0

Note – $\lambda_{\max}^{AR}$ and $\lambda_{\max}^{MA}$ are the respective dominant eigenvalues of $\tilde{A}$ (autoregressive part) and $\tilde{B}$ (moving-average part) as described in (5.18). The simulated true models (characterized by the true Kronecker indices) as well as their corresponding complex-conjugate eigenvalues are marked by an asterisk. The eigenvalues not marked by asterisk are real.

Table xii. Frequencies of estimated Kronecker indices for several bivariate ARMA models with different sample sizes (continued)

(p_1, p_2)	$\lambda_{\max}^{AR}$	$\lambda_{\max}^{MA}$	T=100	T=200	T=300	T=400	T=500	T=1000	T=2000
$(1,3)^*$	0.9572*	0.9315	976	997	998	997	1000	1000	1000
$(1,1)$	-	-	1	0	0	0	0	0	0
$(1,2)$	-	-	8	0	0	0	0	0	0
$(1,4)$	-	-	8	2	1	0	0	0	0
$(2,3)$	-	-	7	1	1	3	0	0	0
$(2,3)^*$	0.9829	0.7901	535	764	907	958	985	1000	1000
$(1,3)$	-	-	3	0	0	0	0	0	0
$(2,1)$	-	-	227	16	2	0	0	0	0
$(2,2)$	-	-	218	209	80	32	8	0	0
$(2,4)$	-	-	10	10	11	10	7	0	0
$(3,2)$	-	-	3	1	0	0	0	0	0
$(3,3)$	-	-	4	0	0	0	0	0	0
$(3,3)^*$	0.7972*	0.8286*	566	817	799	833	864	963	998
$(2,1)$	-	-	3	0	0	0	0	0	0
$(2,2)$	-	-	9	0	0	0	0	0	0
$(2,3)$	-	-	156	16	7	1	0	0	0
$(2,4)$	-	-	2	0	1	0	0	0	0
$(3,1)$	-	-	67	3	0	0	0	0	0
$(3,2)$	-	-	184	151	179	159	128	36	1
$(3,4)$	-	-	13	12	12	3	4	0	0
$(4,3)$	-	-	0	1	2	4	4	1	1
$(3,2)^*$	0.8147*	0.8900*	683	923	980	995	998	1000	1000
$(2,2)$	-	-	269	59	15	4	0	0	0
$(2,3)$	-	-	2	0	0	0	0	0	0
$(3,3)$	-	-	41	17	5	1	2	0	0
$(3,4)$	-	-	3	1	0	0	0	0	0
$(4,2)$	-	-	2	0	0	0	0	0	0
$(3,1)^*$	0.9720*	0.6298*	947	989	998	1000	1000	1000	1000
$(2,1)$	-	-	34	8	1	0	0	0	0
$(3,2)$	-	-	18	3	1	0	0	0	0
$(3,3)$	-	-	1	0	0	0	0	0	0
$(3,0)^*$	0.9380*	0.7040	757	914	942	977	987	998	1000
$(2,1)$	-	-	3	0	0	0	0	0	0
$(2,2)$	-	-	4	0	0	0	0	0	0
$(3,1)$	-	-	194	66	51	19	11	2	0
$(3,2)$	-	-	19	10	2	0	0	0	0
$(4,0)$	-	-	16	10	4	4	2	0	0
$(5,0)$	-	-	7	0	1	0	0	0	0

Note – $\lambda_{\max}^{AR}$ and $\lambda_{\max}^{MA}$ are the respective dominant eigenvalues of $\tilde{A}$ (autoregressive part) and $\tilde{B}$ (moving-average part) as described in (5.18). The simulated true models (characterized by the true Kronecker indices) as well as their corresponding complex-conjugate eigenvalues are marked by an asterisk. The eigenvalues not marked by asterisk are real.

Table xiii. VARMA model with Kronecker indices (1,1) and sample of 200 observations

Second Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	0.0002	0.0002	0.0058	-0.1895	0.1701	0.1076	0.1076
$\mu_{\Phi,2}$	0.0000	-0.0000	0.0000	0.0001	-0.0926	0.0894	0.0566	0.0566
$\phi_{11,1}$	1.2000	1.2039	0.0039	1.2083	1.1340	1.2622	0.0392	0.0390
$\phi_{21,1}$	0.3000	0.3082	0.0082	0.3067	0.2698	0.3480	0.0260	0.0247
$\phi_{12,1}$	-0.4000	-0.4237	0.0237	-0.4292	-0.5180	-0.3130	0.0670	0.0626
$\phi_{22,1}$	0.5000	0.4924	0.0075	0.4958	0.4203	0.5547	0.0411	0.0404
$\theta_{11,1}$	0.3400	0.3150	0.0249	0.3141	0.1797	0.4547	0.0905	0.0869
$\theta_{21,1}$	0.4200	0.4134	0.0065	0.4135	0.3057	0.5157	0.0640	0.0637
$\theta_{12,1}$	-0.6000	-0.5739	0.0260	-0.5771	-0.7615	-0.3786	0.1204	0.1176
$\theta_{22,1}$	0.3000	0.2873	0.0126	0.2849	0.1402	0.4404	0.0927	0.0919

Third Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	0.0004	0.0004	0.0068	-0.1883	0.1771	0.1087	0.1087
$\mu_{\Phi,2}$	0.0000	0.0000	0.0000	-0.0000	-0.0896	0.0889	0.0562	0.0562
$\phi_{11,1}$	1.2000	1.1927	0.0072	1.1956	1.1221	1.2495	0.0398	0.0391
$\phi_{21,1}$	0.3000	0.3070	0.0070	0.3071	0.2694	0.3457	0.0246	0.0235
$\phi_{12,1}$	-0.4000	-0.4052	0.0052	-0.4108	-0.4975	-0.2941	0.0626	0.0624
$\phi_{22,1}$	0.5000	0.4940	0.0059	0.4956	0.4247	0.5529	0.0390	0.0386
$\theta_{11,1}$	0.3400	0.3387	0.0012	0.3383	0.2270	0.4486	0.0662	0.0662
$\theta_{21,1}$	0.4200	0.4115	0.0084	0.4111	0.3239	0.4990	0.0539	0.0532
$\theta_{12,1}$	-0.6000	-0.5954	0.0045	-0.5964	-0.7281	-0.4595	0.0827	0.0825
$\theta_{22,1}$	0.3000	0.2931	0.0068	0.2931	0.1734	0.4145	0.0738	0.0735

Note – These estimates are obtained from the estimated frequencies of the true model over 1000 replications [see table xi]. Note also that the numbers of eigenvalues (the invert of the characteristic roots of the model) associated with the matrices $\tilde{A}$ (autoregressive part) and $\tilde{B}$ (moving-average part) as described in (5.18) are each equal to the McMillan degree $\sum_{l=1}^{k} p_l$. So stationary invertible VARMA models imply that such eigenvalues must be inside the unit circle. For the present model these eigenvalues are real 0.9000 and 0.8000 for the autoregressive part, and conjugate -0.3200$\pm$0.5015i (0.5949 in norm) for the moving-average part.

Table xiv. VARMA model with Kronecker indices (1,2) and sample of 300 observations

Second Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	0.0017	0.0017	-0.0005	-0.1259	0.1340	0.0821	0.0821
$\mu_{\Phi,2}$	0.0000	-0.0023	0.0023	-0.0030	-0.1160	0.1107	0.0684	0.0684
$\phi_{11,1}$	1.2000	1.1988	0.0011	1.1998	1.1519	1.2436	0.0277	0.0276
$\phi_{12,1}$	0.2400	0.2446	0.0046	0.2447	0.2005	0.2842	0.0256	0.0252
$\phi_{22,1}$	0.4000	0.4218	0.0218	0.4254	0.3170	0.5217	0.0673	0.0636
$\phi_{21,2}$	-0.9000	-0.8884	0.0115	-0.8873	-0.9657	-0.8178	0.0462	0.0448
$\phi_{22,2}$	-0.2700	-0.2777	0.0077	-0.2783	-0.3367	-0.2179	0.0380	0.0372
$\theta_{11,1}$	0.8000	0.7817	0.0182	0.7812	0.6795	0.8885	0.0662	0.0637
$\theta_{21,1}$	0.5000	0.4980	0.0019	0.4976	0.4156	0.5809	0.0511	0.0511
$\theta_{12,1}$	0.4000	0.3922	0.0077	0.3938	0.2503	0.5332	0.0844	0.0840
$\theta_{22,1}$	0.4000	0.3631	0.0368	0.3633	0.2398	0.4791	0.0826	0.0740
$\theta_{21,2}$	0.3400	0.3033	0.0366	0.3042	0.1527	0.4472	0.0971	0.0899
$\theta_{22,2}$	0.8500	0.8162	0.0337	0.8152	0.6916	0.9443	0.0864	0.0795

Third Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	0.0018	0.0018	0.0010	-0.1343	0.1348	0.0839	0.0838
$\mu_{\Phi,2}$	0.0000	-0.0022	0.0022	-0.0021	-0.1174	0.1140	0.0711	0.0711
$\phi_{11,1}$	1.2000	1.1943	0.0056	1.1963	1.1463	1.2383	0.0283	0.0277
$\phi_{12,1}$	0.2400	0.2408	0.0008	0.2409	0.1976	0.2808	0.0250	0.0250
$\phi_{22,1}$	0.4000	0.3985	0.0014	0.3954	0.3259	0.4810	0.0460	0.0460
$\phi_{21,2}$	-0.9000	-0.9037	0.0037	-0.9041	-0.9632	-0.8461	0.0359	0.0357
$\phi_{22,2}$	-0.2700	-0.2683	0.0016	-0.2687	-0.3190	-0.2172	0.0308	0.0307
$\theta_{11,1}$	0.8000	0.8038	0.0038	0.8035	0.7178	0.8869	0.0503	0.0502
$\theta_{21,1}$	0.5000	0.4992	0.0007	0.4991	0.4166	0.5778	0.0483	0.0483
$\theta_{12,1}$	0.4000	0.4036	0.0036	0.4031	0.3036	0.4981	0.0589	0.0588
$\theta_{22,1}$	0.4000	0.3948	0.0051	0.3968	0.3100	0.4813	0.0521	0.0518
$\theta_{21,2}$	0.3400	0.3364	0.0035	0.3395	0.2158	0.4478	0.0725	0.0724
$\theta_{22,2}$	0.8500	0.8409	0.0090	0.8399	0.7369	0.9447	0.0651	0.0645

Note – These estimates are obtained from the estimated frequencies of the true model over 1000 replications [see table xi]. The eigenvalues of the model are real 0.9000, 0.4000 and 0.3000 for the autoregressive part, and real -0.8240 and conjugate $-0.1879 \pm 0.7904i$ (0.8125 in norm) for the moving-average part.

Table xv. VARMA model with Kronecker indices (3,1) and sample of 300 observations

Second Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	-0.0009	0.0009	-0.0004	-0.1149	0.1104	0.0679	0.0679
$\mu_{\Phi,2}$	0.0000	0.0006	0.0006	0.0009	-0.0781	0.0785	0.0467	0.0467
$\phi_{21,0}$	0.5000	0.4997	0.0002	0.4969	0.4562	0.5550	0.0301	0.0301
$\phi_{11,1}$	0.4000	0.4065	0.0065	0.4015	0.2116	0.6198	0.1225	0.1224
$\phi_{21,1}$	-0.7000	-0.6964	0.0035	-0.6961	-0.7423	-0.6565	0.0262	0.0260
$\phi_{22,1}$	0.7500	0.7504	0.0004	0.7518	0.7067	0.7888	0.0260	0.0260
$\phi_{11,2}$	0.3500	0.3441	0.0058	0.3482	0.2223	0.4556	0.0748	0.0745
$\phi_{11,3}$	-0.6000	-0.6003	0.0003	-0.6031	-0.6633	-0.5288	0.0463	0.0463
$\phi_{12,3}$	0.6500	0.6436	0.0063	0.6435	0.4646	0.8138	0.1074	0.1072
$\theta_{11,1}$	-0.1000	-0.1161	0.0161	-0.1198	-0.3312	0.1095	0.1377	0.1367
$\theta_{21,1}$	0.3000	0.2978	0.0021	0.2988	0.1818	0.4165	0.0707	0.0707
$\theta_{12,1}$	0.2000	0.1987	0.0012	0.1981	0.0550	0.3381	0.0845	0.0845
$\theta_{22,1}$	0.4000	0.3833	0.0166	0.3790	0.2408	0.5346	0.0902	0.0887
$\theta_{11,2}$	0.6400	0.6316	0.0083	0.6344	0.5140	0.7444	0.0719	0.0715
$\theta_{12,2}$	-0.5800	-0.5878	0.0078	-0.5879	-0.7627	-0.4265	0.1050	0.1047
$\theta_{11,3}$	-0.2700	-0.2702	0.0002	-0.2766	-0.5021	-0.0145	0.1468	0.1468
$\theta_{12,3}$	-0.4600	-0.4232	0.0367	-0.4391	-0.7569	-0.0225	0.2267	0.2237

Third Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	-0.0008	0.0008	-0.0007	-0.1146	0.1141	0.0685	0.0685
$\mu_{\Phi,2}$	0.0000	0.0007	0.0007	0.0012	-0.0778	0.0775	0.0474	0.0474
$\phi_{21,0}$	0.5000	0.5024	0.0024	0.4994	0.4561	0.5605	0.0313	0.0312
$\phi_{11,1}$	0.4000	0.4001	0.0001	0.4039	0.2789	0.5146	0.0739	0.0739
$\phi_{21,1}$	-0.7000	-0.6978	0.0021	-0.6966	-0.7443	-0.6553	0.0270	0.0269
$\phi_{22,1}$	0.7500	0.7485	0.0014	0.7496	0.7027	0.7892	0.0263	0.0262
$\phi_{11,2}$	0.3500	0.3478	0.0021	0.3513	0.2461	0.4307	0.0589	0.0588
$\phi_{11,3}$	-0.6000	-0.6022	0.0022	-0.6033	-0.6609	-0.5379	0.0443	0.0443
$\phi_{12,3}$	0.6500	0.6494	0.0005	0.6481	0.5439	0.7567	0.0694	0.0694
$\theta_{11,1}$	-0.1000	-0.1084	0.0084	-0.1158	-0.2616	0.0724	0.1014	0.1010
$\theta_{21,1}$	0.3000	0.3001	0.0001	0.3013	0.1725	0.4254	0.0730	0.0730
$\theta_{12,1}$	0.2000	0.2005	0.0005	0.2009	0.0720	0.3407	0.0843	0.0843
$\theta_{22,1}$	0.4000	0.3922	0.0077	0.3893	0.2436	0.5432	0.0909	0.0905
$\theta_{11,2}$	0.6400	0.6320	0.0079	0.6342	0.5405	0.7220	0.0578	0.0573
$\theta_{12,2}$	-0.5800	-0.5886	0.0086	-0.5873	-0.7156	-0.4697	0.0815	0.0810
$\theta_{11,3}$	-0.2700	-0.2755	0.0055	-0.2841	-0.4558	-0.0643	0.1176	0.1174
$\theta_{12,3}$	-0.4600	-0.4561	0.0038	-0.4607	-0.6509	-0.2400	0.1260	0.1259

Note – These estimates are obtained from the estimated frequencies of the true model over 1000 replications [see table xii]. The eigenvalues of the model are real -0.5692, -0.0092 and conjugate $0.8642^{+}_{-}0.4447$i (0.9720 in norm) for the autoregressive part, and conjugate $-0.3560^{+}_{-}0.5195$i (0.6298 in norm) and $0.1560^{+}_{-}0.2263$i (0.2749 in norm) for the moving-average part.

Table xvi. VARMA model with Kronecker indices (2,1) and sample of 500 observations

Second Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	-0.0000	0.0000	0.0003	-0.0633	0.0617	0.0395	0.0395
$\mu_{\Phi,2}$	0.0000	0.0000	0.0000	-0.0009	-0.0645	0.0602	0.0389	0.0389
$\phi_{21,0}$	0.5000	0.4966	0.0033	0.4971	0.4736	0.5189	0.0142	0.0138
$\phi_{11,1}$	1.8000	1.8053	0.0053	1.8054	1.7803	1.8304	0.0165	0.0156
$\phi_{21,1}$	-0.4000	-0.3847	0.0152	-0.3871	-0.4530	-0.3109	0.0454	0.0427
$\phi_{22,1}$	0.8000	0.7767	0.0232	0.7793	0.6652	0.8716	0.0664	0.0622
$\phi_{11,2}$	-0.3600	-0.3832	0.0232	-0.3844	-0.4649	-0.2973	0.0555	0.0504
$\phi_{12,2}$	-0.9000	-0.8643	0.0356	-0.8619	-0.9915	-0.7446	0.0828	0.0747
$\theta_{11,1}$	0.3300	0.3182	0.0117	0.3194	0.2374	0.4009	0.0519	0.0506
$\theta_{21,1}$	-0.1800	-0.1849	0.0049	-0.1858	-0.2753	-0.0909	0.0565	0.0563
$\theta_{12,1}$	-0.2000	-0.2004	0.0004	-0.1972	-0.3107	-0.0928	0.0644	0.0644
$\theta_{22,1}$	-0.4000	-0.3823	0.0176	-0.3833	-0.5406	-0.2354	0.0942	0.0925
$\theta_{11,2}$	-0.2000	-0.1957	0.0042	-0.1946	-0.2982	-0.0967	0.0606	0.0605
$\theta_{12,2}$	0.9200	0.8854	0.0345	0.8850	0.7152	1.0493	0.1071	0.1013

Third Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	-0.0002	0.0002	0.0003	-0.0620	0.0606	0.0377	0.0377
$\mu_{\Phi,2}$	0.0000	0.0002	0.0002	-0.0004	-0.0623	0.0594	0.0376	0.0376
$\phi_{21,0}$	0.5000	0.4996	0.0003	0.4997	0.4797	0.5195	0.0123	0.0123
$\phi_{11,1}$	1.8000	1.7996	0.0003	1.7998	1.7780	1.8221	0.0136	0.0136
$\phi_{21,1}$	-0.4000	-0.3952	0.0047	-0.3962	-0.4503	-0.3339	0.0354	0.0350
$\phi_{22,1}$	0.8000	0.7917	0.0082	0.7944	0.6978	0.8707	0.0513	0.0506
$\phi_{11,2}$	-0.3600	-0.3609	0.0009	-0.3611	-0.4262	-0.2946	0.0415	0.0415
$\phi_{12,2}$	-0.9000	-0.8979	0.0020	-0.8980	-0.9998	-0.8018	0.0619	0.0618
$\theta_{11,1}$	0.3300	0.3209	0.0090	0.3220	0.2473	0.3952	0.0472	0.0463
$\theta_{21,1}$	-0.1800	-0.1772	0.0027	-0.1779	-0.2446	-0.1067	0.0413	0.0412
$\theta_{12,1}$	-0.2000	-0.2039	0.0039	-0.2045	-0.2995	-0.1038	0.0569	0.0567
$\theta_{22,1}$	-0.4000	-0.3912	0.0087	-0.3901	-0.4988	-0.2879	0.0638	0.0632
$\theta_{11,2}$	-0.2000	-0.2080	0.0080	-0.2065	-0.2848	-0.1338	0.0473	0.0467
$\theta_{12,2}$	0.9200	0.9169	0.0030	0.9132	0.7982	1.0465	0.0770	0.0770

Note – These estimates are obtained from the estimated frequencies of the true model over 1000 replications [see table xi]. The eigenvalues of the model are real 0.9000, 0.9000 and 0.8000 for the autoregressive part, and real -0.5301 and conjugate $0.3500^{+}_{-}0.5836$i (0.6806 in norm) for the moving-average part.

Table xvii. VARMA model with Kronecker indices (1,3) and sample of 500 observations

Second Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	0.0000	0.0000	0.0000	-0.0097	0.0091	0.0056	0.0056
$\mu_{\Phi,2}$	0.0000	0.0009	0.0009	0.0008	-0.0928	0.0907	0.0560	0.0560
$\phi_{11,1}$	0.3000	0.2796	0.0203	0.2809	0.2105	0.3439	0.0455	0.0407
$\phi_{12,1}$	0.4500	0.4568	0.0068	0.4557	0.4139	0.5055	0.0292	0.0284
$\phi_{22,1}$	0.5000	0.4890	0.0109	0.4866	0.3769	0.6109	0.0710	0.0702
$\phi_{22,2}$	-0.3700	-0.3691	0.0008	-0.3683	-0.4678	-0.2743	0.0568	0.0568
$\phi_{21,3}$	0.6000	0.5705	0.0294	0.5711	0.4643	0.6646	0.0665	0.0596
$\phi_{22,3}$	-0.4000	-0.4002	0.0002	-0.4032	-0.4832	-0.3057	0.0539	0.0539
$\theta_{11,1}$	-0.9500	-0.8905	0.0594	-0.8940	-0.9719	-0.8018	0.0793	0.0525
$\theta_{21,1}$	0.9000	0.8936	0.0063	0.8934	0.8346	0.9593	0.0387	0.0382
$\theta_{12,1}$	0.2000	0.2179	0.0179	0.2146	0.0909	0.3491	0.0777	0.0756
$\theta_{22,1}$	0.3000	0.3049	0.0049	0.3056	0.1676	0.4390	0.0828	0.0826
$\theta_{21,2}$	-0.5400	-0.4859	0.0540	-0.4857	-0.6353	-0.3454	0.1048	0.0898
$\theta_{22,2}$	0.6000	0.6324	0.0324	0.6356	0.4854	0.7690	0.0922	0.0863
$\theta_{21,3}$	-0.4600	-0.3990	0.0609	-0.4015	-0.5693	-0.2251	0.1212	0.1048
$\theta_{22,3}$	0.3500	0.3733	0.0233	0.3730	0.1983	0.5522	0.1095	0.1069

Third Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	-0.0000	0.0000	0.0000	-0.0073	0.0068	0.0044	0.0044
$\mu_{\Phi,2}$	0.0000	0.0009	0.0009	0.0019	-0.0928	0.0857	0.0544	0.0544
$\phi_{11,1}$	0.3000	0.2983	0.0016	0.2986	0.2492	0.3450	0.0295	0.0295
$\phi_{12,1}$	0.4500	0.4563	0.0063	0.4556	0.4127	0.5038	0.0288	0.0281
$\phi_{22,1}$	0.5000	0.4951	0.0048	0.4973	0.4283	0.5542	0.0379	0.0376
$\phi_{22,2}$	-0.3700	-0.3597	0.0102	-0.3596	-0.4204	-0.3010	0.0381	0.0367
$\phi_{21,3}$	0.6000	0.5916	0.0083	0.5917	0.5134	0.6661	0.0479	0.0472
$\phi_{22,3}$	-0.4000	-0.4025	0.0025	-0.4023	-0.4544	-0.3516	0.0315	0.0314
$\theta_{11,1}$	-0.9500	-0.9479	0.0020	-0.9505	-0.9912	-0.8962	0.0292	0.0291
$\theta_{21,1}$	0.9000	0.9041	0.0041	0.9017	0.8464	0.9631	0.0356	0.0354
$\theta_{12,1}$	0.2000	0.1941	0.0058	0.1928	0.0917	0.2902	0.0602	0.0599
$\theta_{22,1}$	0.3000	0.3052	0.0052	0.3048	0.2156	0.3972	0.0552	0.0550
$\theta_{21,2}$	-0.5400	-0.5313	0.0086	-0.5319	-0.6272	-0.4272	0.0616	0.0610
$\theta_{22,2}$	0.7000	0.5914	0.0085	0.5944	0.4646	0.7104	0.0749	0.0744
$\theta_{21,3}$	-0.4600	-0.4461	0.0138	-0.4447	-0.5787	-0.3191	0.0791	0.0778
$\theta_{22,3}$	0.3500	0.3529	0.0029	0.3521	0.2502	0.4536	0.0601	0.0601

Note – These estimates are obtained from the estimated frequencies of the true model over 1000 replications [see table xii]. The eigenvalues of the model are real -0.6334 and 0.6718 and conjugate $0.3808^{+}_{-}0.8782i$ (0.9572 in norm) for the autoregressive part, and real -0.5056 and 0.9315 and conjugate $0.1120^{+}_{-}0.7057i$ (0.7145 in norm) for the moving-average part.

Table xviii. Frequencies of estimated Kronecker indices of VARMA model with true Kronecker indices (4,3) : Comparison with Hannan-Kavalieris and Poskitt methods for different sample sizes

				Our method		H-K method		Poskitt method	
T	(p_1, p_2)	$\lambda_{\max}^{AR}$	$\lambda_{\max}^{MA}$	M1	M2	M1	M2	M1	M2
	$(4,3)^*$	0.9792*	0.8831	814	985	777	4	229	1
	$(4,4)$	-	-	97	14	201	18	672	47
	$(4,5)$	-	-	89	1	21	275	99	887
200	$(5,5)$	-	-	0	0	0	552	0	62
	$(5 \leq p_1 \leq 8, 6)$	-	-	0	0	0	15	0	1
	$(6 \leq p_1 \leq 8, 5)$	-	-	0	0	0	82	0	2
	$(7 \leq p_1 \leq 8, 7)$	-	-	0	0	0	54	0	0
	$(4,3)^*$	0.9792*	0.8831	963	994	850	5	1	0
	$(4,4)$	-	-	20	4	133	25	889	9
	$(4,5)$	-	-	17	2	17	395	110	944
500	$(5,5)$	-	-	0	0	0	512	0	45
	$(5 \leq p_1 \leq 7, 6)$	-	-	0	0	0	3	0	1
	$(6 \leq p_1 \leq 8, 5)$	-	-	0	0	0	46	0	1
	$(7,7)$	-	-	0	0	0	14	0	0
	$(4,3)^*$	0.9792*	0.8831	990	996	793	2	0	0
	$(4,4)$	-	-	9	4	185	16	826	4
	$(4,5)$	-	-	1	0	22	397	174	940
1000	$(5,5)$	-	-	0	0	0	549	0	54
	$(6 \leq p_1 \leq 8, 5)$	-	-	0	0	0	25	0	2
	$(6,6)$	-	-	0	0	0	1	0	0
	$(7,7)$	-	-	0	0	0	10	0	0
	$(4,3)^*$	0.9792*	0.8831	996	998	822	0	0	0
	$(4,4)$	-	-	3	1	167	11	846	1
	$(4,5)$	-	-	1	1	11	364	154	935
2000	$(5,5)$	-	-	0	0	0	594	0	63
	$(5 \leq p_1 \leq 6, 6)$	-	-	0	0	0	2	0	1
	$(6 \leq p_1 \leq 7, 5)$	-	-	0	0	0	20	0	0
	$(7,7)$	-	-	0	0	0	9	0	0

Note – $\lambda_{\max}^{AR}$ and $\lambda_{\max}^{MA}$ are the respective dominant eigenvalues of $\tilde{A}$ (autoregressive part) and $\tilde{B}$ (moving-average part) as described in (5.18). The simulated true model (characterized by the true Kronecker indices) as well as its corresponding complex-conjugate eigenvalues are marked by an asterisk. The eigenvalues not marked by asterisk are real.

Table xix. VARMA model with Kronecker indices (4,3) and sample of 500 observations

Second Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	0.0005	0.0005	0.0019	-0.0410	0.0418	0.0253	0.0253
$\mu_{\Phi,2}$	0.0000	0.0003	0.0003	0.0002	-0.0131	0.0146	0.0086	0.0086
$\phi_{21,0}$	0.5000	0.5013	0.0013	0.5042	0.3896	0.6228	0.0804	0.0804
$\phi_{11,1}$	1.8500	1.8451	0.0048	1.8457	1.7296	1.9741	0.0823	0.0822
$\phi_{21,1}$	0.9000	0.8774	0.0225	0.8767	0.6969	1.0576	0.1275	0.1255
$\phi_{22,1}$	0.2000	0.2137	0.0137	0.2091	0.0272	0.4007	0.1200	0.1192
$\phi_{11,2}$	-0.2400	-0.2795	0.0395	-0.2665	-0.4819	-0.0898	0.1416	0.1359
$\phi_{21,2}$	-0.5000	-0.4870	0.0129	-0.4916	-0.6912	-0.2824	0.1548	0.1542
$\phi_{12,2}$	-1.9700	-1.9319	0.0380	-1.9269	-2.1335	-1.7487	0.1236	0.1176
$\phi_{22,2}$	1.0000	0.9998	0.0001	1.0036	0.8309	1.1728	0.1290	0.1290
$\phi_{11,3}$	0.8900	0.8844	0.0055	0.8745	0.6789	1.1251	0.1384	0.1383
$\phi_{21,3}$	-0.2000	-0.1986	0.0013	-0.2003	-0.3333	-0.0694	0.1041	0.1041
$\phi_{12,3}$	1.4600	1.4399	0.0200	1.4433	1.2645	1.6417	0.1385	0.1371
$\phi_{22,3}$	0.4000	0.3740	0.0259	0.3798	0.2211	0.5254	0.1717	0.1698
$\phi_{11,4}$	-0.8000	-0.7782	0.0217	-0.7738	-0.9306	-0.6383	0.0919	0.0893
$\phi_{12,4}$	1.0500	1.0012	0.0487	1.0046	0.8466	1.1489	0.1078	0.0961
$\theta_{11,1}$	-0.8500	-0.8084	0.0415	-0.8096	-0.9678	-0.6521	0.1101	0.1020
$\theta_{21,1}$	-1.0300	-0.9983	0.0316	-1.0015	-1.1645	-0.8307	0.1242	0.1201
$\theta_{12,1}$	0.6000	0.6279	0.0279	0.6340	0.5211	0.7365	0.1039	0.1000
$\theta_{22,1}$	-0.2000	-0.2013	0.0013	-0.1975	-0.3717	-0.0399	0.0997	0.0996
$\theta_{11,2}$	-1.0400	-1.0028	0.0371	-1.0069	-1.1732	-0.8448	0.1304	0.1250
$\theta_{21,2}$	0.8900	0.8355	0.0544	0.8410	0.6510	1.0332	0.1446	0.1339
$\theta_{12,2}$	0.8800	0.8346	0.0453	0.8393	0.5755	1.0540	0.1507	0.1438
$\theta_{22,2}$	0.1000	0.0538	0.0461	0.0579	-0.1584	0.2502	0.1392	0.1313
$\theta_{11,3}$	0.5300	0.5229	0.0070	0.5251	0.3065	0.7402	0.1470	0.1469
$\theta_{21,3}$	0.4100	0.3796	0.0303	0.3835	0.1439	0.5984	0.1434	0.1402
$\theta_{12,3}$	0.0600	0.0501	0.0098	0.0549	-0.2102	0.3106	0.1615	0.1612
$\theta_{22,3}$	-1.0000	-0.9787	0.0212	-0.9792	-1.1853	-0.7756	0.1455	0.1439
$\theta_{11,4}$	0.2500	0.1903	0.0596	0.1940	-0.0366	0.4162	0.1551	0.1432
$\theta_{12,4}$	-0.6000	-0.5905	0.0094	-0.5909	-0.8031	-0.3697	0.1486	0.1483

Note – These estimates are obtained from the estimated frequencies of the true model over 1000 replications (see table xviii, M2 of our method). The eigenvalues of the model are real 0.4331, -0.7946 and -0.6031, and conjugate $0.9500 \mp 0.2372i$ (0.9792 in norm) and $0.5573 \mp 0.4919i$ (0.7433 in norm) for the autoregressive part, and real 0.8831, -0.6014 and -0.2304, and conjugate $0.2229 \mp 0.7989i$ (0.8294 in norm) and $0.1263 \mp 0.1775i$ (0.2179 in norm) for the moving-average part.

Table xx. VARMA model with Kronecker indices (4,3) and sample of 500 observations (continued)

Third Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	0.0005	0.0005	0.0008	-0.0429	0.0436	0.0258	0.0258
$\mu_{\Phi,2}$	0.0000	0.0003	0.0003	0.0000	-0.0117	0.0127	0.0081	0.0081
$\phi_{21,0}$	0.5000	0.4965	0.0034	0.4980	0.3843	0.6178	0.0799	0.0798
$\phi_{11,1}$	1.8500	1.8462	0.0037	1.8471	1.7417	1.9679	0.0768	0.0767
$\phi_{21,1}$	0.9000	0.8976	0.0023	0.8959	0.7077	1.0833	0.1278	0.1278
$\phi_{22,1}$	0.2000	0.2029	0.0029	0.1967	0.0258	0.3928	0.1167	0.1167
$\phi_{11,2}$	-0.2400	-0.2520	0.0120	-0.2440	-0.4252	-0.0966	0.1154	0.1147
$\phi_{21,2}$	-0.5000	-0.4908	0.0091	-0.4936	-0.6937	-0.2941	0.1628	0.1626
$\phi_{12,2}$	-1.9700	-1.9712	0.0012	-1.9652	-2.1569	-1.7966	0.1104	0.1104
$\phi_{22,2}$	1.0000	0.9983	0.0016	1.0011	0.8308	1.1687	0.1278	0.1278
$\phi_{11,3}$	0.8900	0.9012	0.0112	0.8921	0.7254	1.1083	0.1205	0.1200
$\phi_{21,3}$	-0.2000	-0.2032	0.0032	-0.2066	-0.3287	-0.0775	0.1088	0.1087
$\phi_{12,3}$	1.4600	1.4563	0.0036	1.4572	1.3027	1.6244	0.1227	0.1227
$\phi_{22,3}$	0.4000	0.3927	0.0072	0.3983	0.2396	0.5474	0.1698	0.1696
$\phi_{11,4}$	-0.8000	-0.8010	0.0010	-0.7965	-0.9223	-0.6893	0.0728	0.0728
$\phi_{12,4}$	1.0500	1.0441	0.0058	1.0481	0.9121	1.1673	0.0828	0.0826
$\theta_{11,1}$	-0.8500	-0.8436	0.0063	-0.8452	-0.9980	-0.6854	0.1011	0.1009
$\theta_{21,1}$	-1.0300	-1.0211	0.0088	-1.0250	-1.1857	-0.8475	0.1194	0.1190
$\theta_{12,1}$	0.6000	0.6008	0.0008	0.6057	0.4935	0.7130	0.0953	0.0953
$\theta_{22,1}$	-0.2000	-0.2040	0.0040	-0.1991	-0.3655	-0.0570	0.0958	0.0957
$\theta_{11,2}$	-1.0400	-1.0342	0.0057	-1.0374	-1.1911	-0.8782	0.1210	0.1209
$\theta_{21,2}$	0.8900	0.8814	0.0085	0.8842	0.7063	1.0619	0.1229	0.1226
$\theta_{12,2}$	0.8800	0.8663	0.0136	0.8738	0.6221	1.0792	0.1393	0.1386
$\theta_{22,2}$	0.1000	0.0884	0.0115	0.0875	-0.1021	0.2783	0.1182	0.1176
$\theta_{11,3}$	0.5300	0.5280	0.0019	0.5321	0.3528	0.7135	0.1281	0.1281
$\theta_{21,3}$	0.4100	0.4112	0.0012	0.4195	0.1834	0.6487	0.1391	0.1391
$\theta_{12,3}$	0.0600	0.0507	0.0092	0.0454	-0.1994	0.3015	0.1519	0.1517
$\theta_{22,3}$	-1.0000	-0.9918	0.0081	-0.9910	-1.1892	-0.7981	0.1444	0.1442
$\theta_{11,4}$	0.2500	0.2393	0.0106	0.2467	0.0459	0.4068	0.1135	0.1130
$\theta_{12,4}$	-0.6000	-0.5974	0.0025	-0.5943	-0.8027	-0.3972	0.1361	0.1361

Note – These estimates are obtained from the estimated frequencies of the true model over 1000 replications (see table xviii, M2 of our method). The eigenvalues of the model are real 0.4331, -0.7946 and -0.6031, and conjugate 0.9500 ∓ 0.2372i (0.9792 in norm) and 0.5573 ∓ 0.4919i (0.7433 in norm) for the autoregressive part, and real 0.8831, -0.6014 and -0.2304, and conjugate 0.2229 ∓ 0.7989i (0.8294 in norm) and 0.1263 ∓ 0.1775i (0.2179 in norm) for the moving-average part.

Table xxi. Frequencies of estimated Kronecker indices of VARMA model with true Kronecker indices (3,1,2) : Comparison with Hannan-Kavalieris and Poskitt methods for different sample sizes

				Our method		H-K method		Poskitt method	
T	(p_1, p_2, p_3)	$\lambda_{\max}^{AR}$	$\lambda_{\max}^{MA}$	M1	M2	M1	M2	M1	M2
	$(0,0,0)$	-	-	0	0	113	0	0	0
	$(0,2,0 \leq p_3 \leq 1)$	-	-	0	0	0	0	6	0
	$(1,0,0)$	-	-	0	0	500	0	0	0
	$(1,1,0)$	-	-	0	0	335	0	9	0
	$(1,1,1)$	-	-	0	0	52	2	465	0
	$(2,1,1)$	-	-	0	0	0	0	326	0
	$(2,1,2)$	-	-	0	1	0	121	27	1
	$(2,2,2)$	-	-	2	1	0	183	14	34
200	$(3,1,1)$	-	-	99	164	0	49	153	4
	$(3,1,2)$*	0.9144	0.8556*	820	822	0	93	1	23
	$(3,1,3)$	-	-	3	0	0	68	0	2
	$(3,2 \leq p_2 \leq 3,1)$	-	-	8	1	0	11	1	7
	$(3,2,2)$	-	-	24	1	0	14	0	333
	$(3,3,2)$	-	-	1	0	0	19	0	563
	$(3,3,3)$	-	-	0	0	0	440	0	31
	$(4 \leq p_1 \leq 5,1,1)$	-	-	12	4	0	0	0	0
	$(4 \leq p_1 \leq 5,1,2)$	-	-	31	6	0	0	0	2
	$(1,1,0)$	-	-	0	0	134	0	0	0
	$(1,1,1)$	-	-	0	0	653	0	0	0
	$(2,1,0)$	-	-	0	0	33	0	0	0
	$(2,1 \leq p_2 \leq 2,1)$	-	-	0	0	4	0	11	0
	$(2,1,2)$	-	-	0	0	171	3	30	0
	$(2,2,2)$	-	-	0	0	5	2	335	1
500	$(3,1,1)$	-	-	2	1	0	1	183	0
	$(3,1,2)$*	0.9144	0.8556*	997	999	0	782	177	0
	$(3,1,3)$	-	-	0	0	0	107	0	0
	$(3,2 \leq p_2 \leq 3,1)$	-	-	0	0	0	0	36	0
	$(3,2,2)$	-	-	1	0	0	17	228	37
	$(3,3,2)$	-	-	0	0	0	18	0	921
	$(3,3,3)$	-	-	0	0	0	70	0	41

Note – $\lambda_{\max}^{AR}$ and $\lambda_{\max}^{MA}$ are the respective dominant eigenvalues of $\tilde{A}$ (autoregressive part) and $\tilde{B}$ (moving-average part) as described in (5.18). The simulated true model (characterized by the true Kronecker indices) as well as its corresponding complex-conjugate eigenvalues are marked by an asterisk. The eigenvalues not marked by asterisk are real.

Table xxii. Frequencies of estimated Kronecker indices of VARMA model with true Kronecker indices (3,1,2) : Comparison with Hannan-Kavalieris and Poskitt methods for different sample sizes (continued)

				Our Method		H-K Method		Poskitt Method	
T	(p_1, p_2, p_3)	$\lambda_{\max}^{AR}$	$\lambda_{\max}^{MA}$	M1	M2	M1	M2	M1	M2
1000	$(2,1,2)$	-	-	0	0	973	0	0	0
	$(2,2,2)$	-	-	0	0	22	0	16	0
	$(3,1,1)$	-	-	0	0	1	0	0	0
	$(3,1,2)$*	0.9144	0.8556*	1000	1000	4	785	2	0
	$(3,1,3)$	-	-	0	0	0	122	0	0
	$(3,2,2)$	-	-	0	0	0	12	870	0
	$(3,3,2)$	-	-	0	0	0	17	112	932
	$(3,3,3)$	-	-	0	0	0	64	0	67
	$(3,4,2)$	-	-	0	0	0	0	0	1
2000	$(2,1,2)$	-	-	0	0	55	0	0	0
	$(2,2,2)$	-	-	0	0	16	0	0	0
	$(3,1,1)$	-	-	0	0	1	0	0	0
	$(3,1,2)$*	0.9144	0.8556*	1000	1000	927	754	0	0
	$(3,1,3)$	-	-	0	0	0	152	0	0
	$(3,2,2)$	-	-	0	0	1	4	88	0
	$(3,3,2)$	-	-	0	0	0	24	912	767
	$(3,3,3)$	-	-	0	0	0	66	0	233

Note – $\lambda_{\max}^{AR}$ and $\lambda_{\max}^{MA}$ are the respective dominant eigenvalues of $\tilde{A}$ (autoregressive part) and $\tilde{B}$ (moving-average part) as described in (5.18). The simulated true model (characterized by the true Kronecker indices) as well as its corresponding complex-conjugate eigenvalues are marked by an asterisk. The eigenvalues not marked by asterisk are real.

Table xxiii. VARMA model with Kronecker indices (3,1,2) and sample of 500 observations

Second Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	-0.0029	0.0029	-0.0014	-0.1140	0.0986	0.0635	0.0635
$\mu_{\Phi,2}$	0.0000	-0.0020	0.0020	-0.0014	-0.0423	0.0344	0.0236	0.0235
$\mu_{\Phi,3}$	0.0000	0.0003	0.0003	-0.0003	-0.0572	0.0579	0.0353	0.0353
$\phi_{21,0}$	-0.5000	-0.4978	0.0021	-0.4977	-0.5631	-0.4304	0.0410	0.0409
$\phi_{31,0}$	0.3000	0.3488	0.0488	0.3434	0.1008	0.6001	0.1620	0.1545
$\phi_{11,1}$	0.4600	0.4618	0.0018	0.4588	0.2966	0.6300	0.1027	0.1027
$\phi_{21,1}$	-0.4300	-0.4382	0.0082	-0.4395	-0.5781	-0.2882	0.0870	0.0866
$\phi_{31,1}$	0.2000	0.2617	0.0617	0.2717	-0.0372	0.5260	0.1924	0.1822
$\phi_{22,1}$	-0.2000	-0.1999	0.0000	-0.2003	-0.3906	-0.0086	0.1143	0.1143
$\phi_{23,1}$	-0.1000	-0.0953	0.0046	-0.0957	-0.1646	-0.0225	0.0445	0.0442
$\phi_{33,1}$	0.3000	0.1792	0.1207	0.1776	-0.0493	0.4274	0.1840	0.1388
$\phi_{11,2}$	0.1900	0.1795	0.0104	0.1774	0.0122	0.3574	0.1051	0.1046
$\phi_{31,2}$	0.3400	0.2760	0.0639	0.2783	-0.0558	0.6115	0.2157	0.2060
$\phi_{32,2}$	-0.1100	-0.2006	0.0906	-0.2046	-0.6527	0.2299	0.2893	0.2747
$\phi_{13,2}$	0.1200	0.1346	0.0146	0.1383	-0.0249	0.2784	0.0972	0.0961
$\phi_{33,2}$	-0.0200	0.0102	0.0302	0.0126	-0.1788	0.1919	0.1184	0.1144
$\phi_{11,3}$	-0.1500	-0.1437	0.0062	-0.1445	-0.3365	0.0519	0.1194	0.1193
$\phi_{12,3}$	0.2800	0.3082	0.0282	0.3081	0.0414	0.5611	0.1582	0.1556
$\phi_{13,3}$	0.3000	0.2967	0.0032	0.2965	0.1732	0.4209	0.0750	0.0749
$\theta_{11,1}$	0.5200	0.5100	0.0099	0.5118	0.3276	0.6800	0.1103	0.1099
$\theta_{21,1}$	0.1500	0.1435	0.0064	0.1465	0.0035	0.2813	0.0811	0.0808
$\theta_{31,1}$	0.2000	0.0899	0.1100	0.0923	-0.1758	0.3484	0.1964	0.1626
$\theta_{12,1}$	0.3700	0.3957	0.0257	0.3935	0.2519	0.5405	0.0942	0.0906
$\theta_{22,1}$	-0.1300	-0.1218	0.0081	-0.1219	-0.3333	0.0807	0.1279	0.1277
$\theta_{32,1}$	0.2600	0.2465	0.0134	0.2416	0.0093	0.4885	0.1563	0.1557
$\theta_{13,1}$	-0.1000	-0.0882	0.0117	-0.0883	-0.1580	-0.0164	0.0443	0.0427
$\theta_{23,1}$	0.2200	0.2205	0.0005	0.2200	0.1329	0.3120	0.0543	0.0543
$\theta_{33,1}$	-0.6000	-0.4770	0.1229	-0.4782	-0.7263	-0.2270	0.1960	0.1526
$\theta_{11,2}$	0.3400	0.3111	0.0288	0.3175	0.0977	0.4973	0.1250	0.1216
$\theta_{31,2}$	-0.1800	-0.1300	0.0499	-0.1277	-0.4071	0.1620	0.1782	0.1711
$\theta_{12,2}$	-0.3200	-0.2980	0.0219	-0.3034	-0.4630	-0.1183	0.1073	0.1050
$\theta_{32,2}$	0.2300	0.3289	0.0989	0.3199	-0.1217	0.8152	0.3097	0.2934
$\theta_{13,2}$	0.1600	0.1577	0.0022	0.1526	0.0011	0.3306	0.1025	0.1025
$\theta_{33,2}$	0.3600	0.2737	0.0862	0.2769	0.0298	0.5099	0.1660	0.1418
$\theta_{11,3}$	-0.1400	-0.1478	0.0078	-0.1510	-0.3337	0.0429	0.1145	0.1142
$\theta_{12,3}$	0.3500	0.3069	0.0430	0.3045	-0.0186	0.6399	0.2114	0.2069
$\theta_{13,3}$	0.1600	0.1629	0.0029	0.1605	-0.0066	0.3260	0.1015	0.1014

Note – These estimates are obtained from the estimated frequencies of the true model over 1000 replications (see table xxi, M2 of our Method). The eigenvalues of the model are real 0.9144 and -0.1433 and conjugate $0.3983 \mp 0.6210i$ (0.7378 in norm) and $-0.5039 \mp 0.2801i$ (0.5765 in norm) for the autoregressive part, and real 0.4392 and -0.2014 and conjugate $-0.2587 \mp 0.8156i$ (0.8556 in norm) and $0.3523 \mp 0.3788i$ (0.5174 in norm) for the moving-average part.

Table xxiv. VARMA model with Kronecker indices (3,1,2) and sample of 500 observations (continued)

Third Stage Estimates

Coefficient	Value	Mean	Avg-Dev	Median	5^{th}-p	95^{th}-p	RMSE	Std-Dev
$\mu_{\Phi,1}$	0.0000	-0.0032	0.0032	-0.0013	-0.1112	0.0991	0.0634	0.0633
$\mu_{\Phi,2}$	0.0000	-0.0022	0.0022	-0.0015	-0.0434	0.0358	0.0241	0.0240
$\mu_{\Phi,3}$	0.0000	0.0000	0.0000	-0.0011	-0.0526	0.0556	0.0327	0.0327
$\phi_{21,0}$	-0.5000	-0.4991	0.0008	-0.4996	-0.5648	-0.4307	0.0413	0.0413
$\phi_{31,0}$	0.3000	0.3044	0.0044	0.3039	0.0940	0.5112	0.1415	0.1414
$\phi_{11,1}$	0.4600	0.4612	0.0012	0.4644	0.3061	0.6128	0.0913	0.0913
$\phi_{21,1}$	-0.4300	-0.4283	0.0016	-0.4311	-0.5745	-0.2725	0.0906	0.0906
$\phi_{31,1}$	0.2000	0.2003	0.0003	0.2119	-0.0748	0.4494	0.1650	0.1650
$\phi_{22,1}$	-0.2000	-0.1917	0.0082	-0.1911	-0.3895	0.0057	0.1190	0.1187
$\phi_{23,1}$	-0.1000	-0.0987	0.0012	-0.0984	-0.1711	-0.0236	0.0454	0.0454
$\phi_{33,1}$	0.3000	0.2870	0.0129	0.2852	0.0843	0.4998	0.1263	0.1256
$\phi_{11,2}$	0.1900	0.1969	0.0069	0.1977	0.0471	0.3440	0.0918	0.0915
$\phi_{31,2}$	0.3400	0.3311	0.0088	0.3409	0.0000	0.6194	0.1900	0.1898
$\phi_{32,2}$	-0.1100	-0.1156	0.0056	-0.1127	-0.4695	0.2692	0.2238	0.2238
$\phi_{13,2}$	0.1200	0.1082	0.0117	0.1106	-0.0307	0.2409	0.0870	0.0862
$\phi_{33,2}$	-0.0200	-0.0136	0.0063	-0.0102	-0.1635	0.1221	0.0857	0.0855
$\phi_{11,3}$	-0.1500	-0.1618	0.0118	-0.1615	-0.3239	0.0013	0.1013	0.1006
$\phi_{12,3}$	0.2800	0.2807	0.0007	0.2863	0.0693	0.4723	0.1250	0.1250
$\phi_{13,3}$	0.3000	0.3070	0.0070	0.3058	0.2169	0.4006	0.0557	0.0553
$\theta_{11,1}$	0.5200	0.5184	0.0015	0.5151	0.3522	0.6790	0.1003	0.1003
$\theta_{21,1}$	0.1500	0.1462	0.0037	0.1482	0.0032	0.2817	0.0835	0.0834
$\theta_{31,1}$	0.2000	0.1957	0.0042	0.2006	-0.0518	0.4351	0.1439	0.1438
$\theta_{12,1}$	0.3700	0.3764	0.0064	0.3754	0.2365	0.5204	0.0843	0.0841
$\theta_{22,1}$	-0.1300	-0.1426	0.0126	-0.1461	-0.3430	0.0775	0.1308	0.1302
$\theta_{32,1}$	0.2600	0.2791	0.0191	0.2741	0.0546	0.4920	0.1504	0.1492
$\theta_{13,1}$	-0.1000	-0.0960	0.0039	-0.0968	-0.1644	-0.0307	0.0401	0.0399
$\theta_{23,1}$	0.2200	0.2172	0.0027	0.2190	0.1257	0.3089	0.0568	0.0568
$\theta_{33,1}$	-0.6000	-0.5825	0.0174	-0.5832	-0.8151	-0.3505	0.1435	0.1424
$\theta_{11,2}$	0.3400	0.3256	0.0143	0.3325	0.1548	0.4952	0.1051	0.1041
$\theta_{31,2}$	-0.1800	-0.1736	0.0063	-0.1754	-0.3845	0.0437	0.1352	0.1351
$\theta_{12,2}$	-0.3200	-0.3221	0.0021	-0.3265	-0.4664	-0.1708	0.0900	0.0900
$\theta_{32,2}$	0.2300	0.2366	0.0066	0.2437	-0.1302	0.6001	0.2269	0.2268
$\theta_{13,2}$	0.1600	0.1720	0.0120	0.1728	0.0225	0.3244	0.0948	0.0941
$\theta_{33,2}$	0.3600	0.3463	0.0136	0.3437	0.1565	0.5304	0.1174	0.1166
$\theta_{11,3}$	-0.1400	-0.1347	0.0052	-0.1360	-0.2677	0.0016	0.0825	0.0823
$\theta_{12,3}$	0.3500	0.3550	0.0050	0.3579	0.1033	0.6130	0.1556	0.1555
$\theta_{13,3}$	0.1600	0.1528	0.0071	0.1531	0.0240	0.2813	0.0784	0.0781

Note – These estimates are obtained from the estimated frequencies of the true model over 1000 replications (see table xxii, M2 of our Method). The eigenvalues of the model are real 0.9144 and -0.1433 and conjugate $0.3983\bar{+}0.6210i$ (0.7378 in norm) and $-0.5039\bar{+}0.2801i$ (0.5765 in norm) for the autoregressive part, and real 0.4392 and -0.2014 and conjugate $-0.2587\bar{+}0.8156i$ (0.8556 in norm) and $0.3523\bar{+}0.3788i$ (0.5174 in norm) for the moving-average part.

Table xxv. Frequencies of estimated Kronecker indices of VARMA model with true Kronecker indices (3,2,4,1), $\lambda_{\max}^{AR} = 0.7968^*$ and $\lambda_{\max}^{MA} = 0.9793^*$: Comparison with Hannan-Kavalieris and Poskitt methods for different sample sizes

		Our method		H-K method		Poskitt method	
T	(p_1, p_2, p_3, p_4)	M1	M2	M1	M2	M1	M2
500	$(2 \leq . \leq 3, 2 \leq . \leq 5, 2, 1)$	16	13	16	0	655	5
	$(2 \leq . \leq 3, 2, 2, 2)$	4	14	22	0	0	0
	$(3, 2, 3, 0 \leq . \leq 3)$	151	157	918	200	341	27
	$(3, 2, 4, 1)^*$	744	737	0	12	3	443
	$(3, 2, 4 \leq . \leq 5, 2 \leq . \leq 4)$	54	48	0	48	0	206
	$(3, 3, 1 \leq . \leq 4, 1 \leq . \leq 4)$	28	24	44	293	1	222
	$(3, 4, 2 \leq . \leq 5, 1 \leq . \leq 4)$	3	6	0	285	0	81
	$(4, 2 \leq . \leq 4, 4, 1 \leq . \leq 4)$	0	1	0	129	0	16
	$(4 \leq . \leq 5, 5, 5, 5)$	0	0	0	7	0	0
	$(5 \leq . \leq 6, 6, 6, 6)$	0	0	0	26	0	0
1000	$(3, 2, 3, 1 \leq . \leq 3)$	56	57	989	61	862	0
	$(3, 2, 4, 1)^*$	888	904	0	55	138	129
	$(3, 2, 4 \leq . \leq 5, 2 \leq . \leq 4)$	35	20	0	75	0	585
	$(3, 3, 3 \leq . \leq 4, 1 \leq . \leq 4)$	19	18	11	223	0	208
	$(3, 4, 3 \leq . \leq 4, 1 \leq . \leq 4)$	2	1	0	439	0	76
	$(4, 2, 4, 2 \leq . \leq 4)$	0	0	0	7	0	1
	$(4, 3 \leq . \leq 4, 4, 1 \leq . \leq 4)$	0	0	0	126	0	1
	$(5 \leq . \leq 6, 6, 6, 6)$	0	0	0	4	0	0
2000	$(3, 2, 3, 1)$	12	9	882	0	262	0
	$(3, 2, 3, 2 \leq . \leq 3)$	0	0	100	3	0	0
	$(3, 2, 4, 1)^*$	974	973	0	28	738	0
	$(3, 2, 4 \leq . \leq 5, 2 \leq . \leq 4)$	7	10	0	56	0	428
	$(3, 3 \leq . \leq 4, 4, 1 \leq . \leq 4)$	7	8	18	762	0	572
	$(3 \leq . \leq 6, 6, 6, 6)$	0	0	0	13	0	0
	$(4, 2, 4, 1 \leq . \leq 4)$	0	0	0	4	0	0
	$(4, 3 \leq . \leq 4, 4, 1 \leq . \leq 4)$	0	0	0	134	0	0

Note – $\lambda_{\max}^{AR}$ and $\lambda_{\max}^{MA}$ are the respective dominant eigenvalues of $\tilde{A}$ (autoregressive part) and $\tilde{B}$ (moving-average part) as described in (5.18). The simulated true model (characterized by the true Kronecker indices) as well as its corresponding complex-conjugate eigenvalues are marked by an asterisk. The eigenvalues not marked by asterisk are real.

Table xxvi. Frequencies of estimated Kronecker indices of VARMA model with true Kronecker indices (5,1,3,2,1), $\lambda_{\max}^{AR} = 0.8524^*$ and $\lambda_{\max}^{MA} = 0.8788^*$: Comparison with Hannan-Kavalieris and Poskitt methods for different sample sizes

		Our method		H-K method		Poskitt method	
T	$(p_1, p_2, p_3, p_4, p_5)$	M1	M2	M1	M2	M1	M2
1000	$(3, 1 \leq . \leq 3, 2 \leq . \leq 3, 2 \leq . \leq 3, 1 \leq . \leq 3)$	0	1	1000	0	137	0
	$(4, 1, 3, 2, 1 \leq . \leq 2)$	4	9	0	0	161	0
	$(4, 2 \leq . \leq 4, 3 \leq . \leq 4, 2 \leq . \leq 4, 2 \leq . \leq 4)$	4	0	0	18	0	9
	$(4 \leq . \leq 5, 1, 2, 2, 2)$	2	0	0	0	446	0
	$(5, 1, 3, 2, 1)^*$	903	920	0	0	0	0
	$(5, 1 \leq . \leq 2, 3, 2 \leq . \leq 3, 2)$	45	36	0	0	256	27
	$(5, 1 \leq . \leq 4, 5, 5, 5)$	0	0	0	143	0	0
	$(5, 2 \leq . \leq 3, 3, 2, 1 \leq . \leq 3)$	22	19	0	0	0	607
	$(5, 3, 3 \leq . \leq 4, 3 \leq . \leq 5, 2 \leq . \leq 5)$	0	0	0	1	0	218
	$(5, 4 \leq . \leq 5, 4 \leq . \leq 5, 4 \leq . \leq 5, 5)$	0	0	0	838	0	0
	$(6 \leq . \leq 7, 1, 3, 2, 1)$	19	15	0	0	0	0
	$(6 \leq . \leq 8, 2 \leq . \leq 3, 3, 2 \leq . \leq 3, 2 \leq . \leq 3)$	1	0	0	0	0	139
2000	$(3, 1 \leq . \leq 3, 3, 2 \leq . \leq 3, 1 \leq . \leq 3)$	0	0	955	0	38	0
	$(4, 1, 3, 2, 1)$	0	0	7	0	0	0
	$(4, 1 \leq . \leq 2, 3, 2, 2)$	0	0	0	0	68	0
	$(4, 1, 3 \leq . \leq 4, 2 \leq . \leq 4, 1)$	0	0	38	0	0	0
	$(5, 1 \leq . \leq 2, 2, 2, 2)$	0	0	0	0	28	0
	$(5, 1, 3, 2, 1)^*$	966	969	0	0	0	0
	$(5, 1, 3, 2, 2)$	25	28	0	0	807	0
	$(5, 2, 3, 2 \leq . \leq 4, 2)$	4	0	0	0	59	39
	$(5, 2 \leq . \leq 3, 5, 5, 5)$	0	0	0	41	0	0
	$(5, 3, 3, 2, 2 \leq . \leq 3)$	0	0	0	0	0	87
	$(5, 3, 3, 3, 2 \leq . \leq 3)$	0	0	0	0	0	698
	$(5, 3 \leq . \leq 4, 3, 2 \leq . \leq 5, 2)$	0	0	0	0	0	43
	$(5, 3 \leq . \leq 4, 4 \leq . \leq 5, 5, 5)$	0	0	0	171	0	0
	$(5, 5, 5, 5, 5)$	0	0	0	788	0	0
	$(6 \leq . \leq 7, 1, 3, 2, 1)$	5	3	0	0	0	0
	$(6 \leq . \leq 9, 2 \leq . \leq 4, 3, 2 \leq . \leq 3, 2 \leq . \leq 3)$	0	0	0	0	0	133

Note – $\lambda_{\max}^{AR}$ and $\lambda_{\max}^{MA}$ are the respective dominant eigenvalues of $\tilde{A}$ (autoregressive part) and $\tilde{B}$ (moving-average part) as described in (5.18). The simulated true model (characterized by the true Kronecker indices) as well as their corresponding complex-conjugate eigenvalues are marked by an asterisk. The eigenvalues not marked by asterisk are real.

Références

AHLFORS, L. V. (1966) : *Complex Analysis : An Introduction to the Theory of Analytic Functions of One Complex Variable*, International Series in Pure and Applied Mathematics. McGraw-Hill, New York, second edn.

AKAIKE, H. (1969) : "Fitting Autoregressive Models for Prediction," *Annals of the Institute of Statistical Mathematics*, 21, 243–247.

——— (1973) : "Information Theory and an Extension of the Maximum Likelihood Principle," in *2nd International Symposium on Information Theory*, ed. by B. N. Petrov, and F. Csaki, pp. 267–281. Akademiai Kiado, Budapest.

——— (1974a) : "Markovian Representations of Stochastic Processes and its Application to the Analysis of Autoregressive Moving Average Processes," *Annals of the Institute of Statistical Mathematics*, 26, 363–387.

——— (1974b) : "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716–723.

AKAIKE, H. (1975) : "Markovian Representation of Stochastic Processes by Canonical Variables," *SIAM Journal on Control and Optimization*, 13, 162–173.

AKAIKE, H. (1976) : "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion," in *System Identification : Advances in Case Studies*, ed. by R. K. Mehra, and D. G. Lainiotis, pp. 27–96. Academic Press, New York.

——— (1977) : "On Entropy Maximization Principle," in *Applications of Statistics*, ed. by P. R. Krishnaiah, pp. 27–41, Amsterdam. North Holland.

ALLENDE, H., AND S. HEILER (1992) : "Recursive Generalized M-Estimates for Autoregressive Moving-Average Models," *Journal of Time Series Analysis*, 13(1), 1–18.

ANDERSON, T. W. (1971) : *The Statistical Analysis of Time Series.* John Wiley & Sons, New York.

ANDREWS, D. W. K. (2000) : "Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space," *Econometrica*, 68, 399–405.

ATHREYA, K. B. (1987) : "Bootstrap of the Mean in the Infinite Variance Case," *The Annals of Statistics*, 15, 724–731.

BALTAGI, B. (ed.) (2001) : *Companion to Theoretical Econometrics*, Blackwell Companions to Contemporary Economics. Basil Blackwell, Oxford, U.K.

BARNARD, G. A. (1963) : "Comment on 'The Spectral Analysis of Point Processes' by M. S. Bartlett," *Journal of the Royal Statistical Society, Series B*, 25, 294.

BARTEL, H., AND H. LÜTKEPOHL (1998) : "Estimating the Kronecker Indices of Cointegrated Echelon-Form VARMA Models," *Econometrics Journal*, 1, C76–C99.

BASAWA, I. V., A. K. MALLIK, W. P. MCCORMICK, J. H. REEVES, AND R. L. TAYLOR (1991) : "Bootstrapping Unstable First-Order Autoregressive Processes," *The Annals of Statistics*, 19, 1098–1101.

BENKWITZ, A., H. LÜTKEPOHL, AND M. H. NEUMANN (2000) : "Problems Related to Confidence Intervals for Impulse Responses of Autoregressive Processes," *Econometric Reviews*, 19, 69–103.

BERK, K. N. (1974) : "Consistent Autoregressive Spectral Estimates," *The Annals of Statistics*, 2(3), 489–502.

BERKOWITZ, J., AND L. KILIAN (2000) : "Recent Developments in Bootstrapping Time Series," *Econometric Reviews*, 19, 1–48, With Comment by Russell Davidson, 49-54.

BERNANKE, B. S., AND A. S. BLINDER (1992) : "The Federal Funds Rate and the Channel of Monetary Transmission," *American Economic Review*, 82, 901–921.

BERNANKE, B. S., AND I. MIHOV (1998) : "Measuring Monetary Policy," *The Quarterly Journal of Economics*, 113(3), 869–902.

BIRNBAUM, Z. W. (1974) : "Computers and Unconventional Test-Statistics," in *Reliability and Biometry*, ed. by F. Proschan, and R. J. Serfling, pp. 441–458. SIAM, Philadelphia, PA.

BLOUGH, S. R. (1992) : "The Relationship between Power and Level for Generic Unit Root Tests in Finite Samples," *Journal of Applied Econometrics*, 7, 295–308.

BOUDJELLABA, H., J.-M. DUFOUR, AND R. ROY (1992) : "Testing Causality Between Two Vectors in Multivariate ARMA Models," *Journal of the American Statistical Association*, 87(420), 1082–1090.

——— (1994) : "Simplified Conditions for Non-Causality between Two Vectors in Multivariate ARMA Models," *Journal of Econometrics*, 63, 271–287.

BOX, G. E. P., AND G. M. JENKINS (1976) : *Time Series Analysis : Forecasting and Control.* Holden-Day, San Francisco, second edn.

CANER, M., AND L. KILIAN (1999) : "Size Distortions of Tests of the Null Hypothesis of Stationarity : Evidence and Implications for the PPP Debate," Technical report, Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220, ikilian@umich.edu.

CHOI, I. (1993) : "Asymptotic Normality of the Least-Squares Estimates for Higher Order Autoregressive Integrated Processes with some Applications," *Econometric Theory*, 9, 263–282.

CHRISTIANO, L. J., AND M. EICHENBAUM (1987) : "Temporal Aggregation and Structural Inference in Macroeconomics," *Carnegie-Rochester Conference on Public Policy*, 26, 63–130.

CHRISTIANO, L. J., AND M. EICHENBAUM (1992) : "Liquidity Effects and the Monetary Transmission Mechanism," *American Economic Review*, 82, 346–353.

CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (1994) : "Identification and the Effects of Monetary Policy Shocks," Working Paper 7, Federal Reserve Bank of Chicago.

——— (1996) : "The Effects of Monetary Policy Shocks : Evidence from the Flow of Funds," *Review of Economics and Statistics*, 78(1), 16–34.

——— (1999) : "Monetary Policy Shocks : What Have We Learned and to What End ?," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1, chap. 2, pp. 65–148. Elsevier, 1 edn.

CHUNG, K. L. (2001) : *A Course in Probability Theory.* Academic Press, New York, third edn.

CHURCHILL, R. V., AND J. W. BROWN (1990) : *Complex Variables and Applications.* McGraw-Hill, New York, fifth edn.

CLEMENTS, M. P., AND D. F. HENDRY (eds.) (2002) : *A Companion to Economic Forecasting*, Blackwell Companions to Contemporary Economics. Blackwell, Oxford, U.K.

COCHRANE, J. H. (1991) : "A Critique of the Application of Unit Root Tests," *Journal of Economic Dynamics and Control*, 15, 275–284.

COOPER, D. M., AND E. F. WOOD (1982) : "Identifying Multivariate Time Series Models," *Journal of Time Series Analysis*, 3(3), 153–164.

DE FRUTOS, R. F., AND G. R. SERRANO (2002) : "A Generalized Least Squares Estimation Method for VARMA Models," *Statistics*, 36(4), 303–316.

DEISTLER, M. (1983) : "The Properties of the Parametrization of ARMAX Systems and their Relevance for Structural Estimation and Dynamic Specification.," *Econometrica*, 51, 1187–1208.

DEISTLER, M., AND E. J. HANNAN (1981) : "Some Properties of the Parameterization of ARMA Systems with Unknown Order," *Journal of Multivariate Analysis*, 11, 474–484.

DHRYMES, P. J. (1998) : *Time Series, Unit Roots, and Cointegration.* Academic Press, San Diego, California, USA.

DOLADO, J. J., AND H. LÜTKEPOHL (1996) : "Making Wald Tests Work for Cointegrated VAR Systems," *Econometric Reviews*, 15, 369–386.

DUFOUR, J.-M. (2003) : "Identification, Weak Instruments and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36(4), 767–808.

——— (2006) : "Monte Carlo Tests with Nuisance Parameters : A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics," *Journal of Econometrics*, 133(2), 443–477.

DUFOUR, J.-M., AND T. JOUINI (2005) : "Asymptotic Distribution of a Simple Linear Estimator for VARMA Models in Echelon Form," in *Statistical Modeling and Ana-*

lysis for Complex Data Problems, ed. by P. Duchesne, and B. Rémillard, chap. 11, pp. 209–240. Kluwer/Springer-Verlag, Canada.

——— (2006) : "Finite-Sample Simulation-Based Tests in VAR Models with Applications to Granger Causality Testing," *Journal of Econometrics*, 135(1-2), 229–254.

DUFOUR, J.-M., AND L. KHALAF (2001) : "Monte Carlo Test Methods in Econometrics," in Baltagi (2001), chap. 23, pp. 494–519.

——— (2002) : "Simulation Based Finite and Large Sample Tests in Multivariate Regressions," *Journal of Econometrics*, 111(2), 303–322.

DUFOUR, J.-M., D. PELLETIER, AND É. RENAULT (2006) : "Short Run and Long Run Causality in Time Series : Inference," *Journal of Econometrics*, 132(2), 337–362.

DUFOUR, J.-M., AND E. RENAULT (1998) : "Short-Run and Long-Run Causality in Time Series : Theory," *Econometrica*, 66, 1099–1125.

DURBIN, J. (1960) : "The Fitting of Time Series Models," *Revue de l'Institut International de Statistique*, 28, 233–244.

DWASS, M. (1957) : "Modified Randomization Tests for Nonparametric Hypotheses," *Annals of Mathematical Statistics*, 28, 181–187.

EFRON, B., AND R. J. TIBSHIRANI (1993) : *An Introduction to the Bootstrap*, vol. 57 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, New York.

FAUST, J. (1996) : "Near Observational Equivalence Problems and Theoretical Size Problems with Unit Root Tests," *Econometric Theory*, 12, 724–732.

FAUST, J. (1999) : "Theoretical Confidence Level Problems with Confidence Intervals for the Spectrum of a Time Series," *Econometrica*, 67, 629–637.

FLORES DE FRUTOS, R., AND G. R. SERRANO (2002) : "A Generalized Least Squares Estimation Method for VARMA Models," *Statistics*, 36(4), 303–316.

GALBRAITH, J. W., AND V. ZINDE-WALSH (1994) : "A Simple, Noniterative Estimator for Moving Average Models," *Biometrika*, 81(1), 143–155.

——— (1997) : "On Some Simple, Autoregression-Based Estimation and Identification Techniques for ARMA Models," *Biometrika*, 84(3), 685–696.

GOFFE, W. L., G. D. FERRIER, AND J. ROGERS (1994) : "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60, 65–99.

GORDON, D. B., AND E. M. LEEPER (1994) : "The Dynamic Impacts of Monetary Policy : An Exercice in Tentative Identification," *Journal of Political Economy*, 102(6), 1228–1247.

GOURIÉROUX, C., AND A. MONFORT (1997) : *Time Series and Dynamic Models.* Cambridge University Press, Cambridge, U.K.

GRANGER, C. W. J. (1969) : "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–459.

HALL, P. (1992) : *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.

HAMILTON, J. D. (1994) : *Time Series Analysis.* Princeton University Press, Princeton, New Jersey.

HANNAN, E. J. (1969a) : "The Estimation of Mixed Moving Average Autoregressive Systems," *Biometrika*, 56(3), 579–593.

——— (1969b) : "The Identification of Vector Mixed Autoregressive-Moving Average Systems," *Biometrika*, 57, 223–225.

——— (1970) : *Multiple Time Series.* John Wiley & Sons, New York.

——— (1971) : "The Identification Problem for Multiple Equation Systems with Moving Average Errors," *Econometrica*, 39, 751–766.

——— (1976a) : "The Asymptotic Distribution of Serial Covariances," *The Annals of Statistics*, 4(2), 396–399.

——— (1976b) : "The Identification and Parameterization of ARMAX and State Space Forms," *Econometrica*, 44(4), 713–723.

——— (1979) : "The Statistical Theory of Linear Systems," in *Developments in Statistics*, ed. by P. R. Krishnaiah, vol. 2, pp. 83–121. Academic Press, New York.

——— (1980a) : "The Estimation of the Order of an ARMA Process," *The Annals of Statistics*, 8(5), 1071–1081.

——— (1980b) : "Recursive Estimation Based on ARMA Models," *The Annals of Statistics*, 8(4), 762–777.

——— (1981) : "Estimating the Dimension of a Linear System," *Journal of Multivariate Analysis*, 11(4), 459–473.

HANNAN, E. J., AND M. DEISTLER (1988) : *The Statistical Theory of Linear Systems.* John Wiley & Sons, New York.

HANNAN, E. J., AND L. KAVALIERIS (1984a) : "A Method for Autoregressive-Moving Average Estimation," *Biometrika*, 71(2), 273–280.

——— (1984b) : "Multivariate Linear Time Series Models," *Advances in Applied Probability*, 16, 492–561.

——— (1986) : "Regression, Autoregression Models," *Journal of Time Series Analysis*, 7(1), 27–49.

HANNAN, E. J., L. KAVALIERIS, AND M. MACKISACK (1986) : "Recursive Estimation of Linear Systems," *Biometrika*, 73(1), 119–133.

HANNAN, E. J., AND B. G. QUINN (1979) : "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society, Series B*, 41, 190–195.

HANNAN, E. J., AND J. RISSANEN (1982) : "Recursive Estimation of Mixed Autoregressive-Moving Average Order," *Biometrika*, 69(1), 81–94, Errata 70 (1983), 303.

HANSEN, P. R., AND S. JOHANSEN (1998) : *Workbook on Cointegration.* Oxford University Press, Oxford, U.K.

HATANAKA, M. (1996) : *Time-Series-Based Econometrics*, Advanced Texts in Econometrics. Oxford University Press, Oxford, U.K.

HENDRY, D. F. (1995) : *Dynamic Econometrics.* Oxford University Press, Oxford, U.K.

HILLMER, S. C., AND G. C. TIAO (1979) : "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models," *Journal of the American Statistical Association*, 74(367), 652–660.

HORN, R. G., AND C. A. JOHNSON (1985) : *Matrix Analysis.* Cambridge University Press, Cambridge, U.K.

HUANG, D., AND L. GUO (1990) : "Estimation of Nonstationary ARMAX Models Based on the Hannan-Rissanen Method," *The Annals of Statistics*, 18(4), 1729–1756.

INOUE, A., AND L. KILIAN (2002a) : "Bootstrapping Autoregressive Processs with Possible Unit Roots," *Econometrica*, 70(1), 377–391.

INOUE, A., AND L. KILIAN (2002b) : "Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in $VAR(\infty)$ Models," *International Economic Review*, 43(2), 309–332.

INOUE, A., AND L. KILIAN (2003) : "The Continuity of the Limit Distribution in the Parameter of Interest is Not Essential for the Validity of the Bootstrap," *Econometric Theory*, 19(6), 944–961.

JEONG, J., AND G. S. MADDALA (1993) : "A Perspective on Application of Bootstrap Methods in Econometrics," in *Handbook of Statistics 11 : Econometrics*, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod, pp. 573–610. North-Holland, Amsterdam.

JOHANSEN, S. (1995) : *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Advanced Texts in Econometrics. Oxford University Press, Oxford, U.K.

KAVALIERIS, L., E. J. HANNAN, AND M. SALAU (2003) : "Generalized Least Squares Estimation of ARMA Models," *Journal of Times Series Analysis*, 24(2), 165–172.

KEATING, J. W. (2002) : "Structural Inference with Long-Run Recursive Empirical Models," *Macroeconomic Dynamics*, 6(2), 266–283.

KILIAN, L. (1998a) : "Confidence Intervals for Impulse Responses Under Departures From Normality," *Econometric Reviews*, 17(1), 1–29.

KILIAN, L. (1998b) : "Small-Sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics*, 80(2), 218–230.

KILIAN, L., AND U. DEMIROGLU (1997) : "Residual-Based Bootstrap Tests for Normality in Autoregresssions," Working Papers 14, Department of Economics, University of Michigan, Ann Arbor, Michigan.

——— (2000) : "Residual-Based Tests for Normality in Autoregressions : Asymptotic Theory and Simulation Evidence," *Journal of Business and Economic Statistics*, 18(1), 40–50.

KIM, K.-S., AND W. D. MCMILLIN (2003) : "Estimating the Effects of Monetary Policy Shocks : Does Lag Structure Matter ?," *Applied Economics*, 35(13), 1515–1526.

KOHN, R. (1981) : "A Note on an Alternative Derivation of the Likelihood of an Autoregressive Moving Average Process," *Economics Letters*, 7, 233–236.

KORAY, F., AND W. D. MCMILLIN (1999) : "Monetary Shocks, the Exchange Rate, and the Trade Balance," *Journal of International Money and Finance*, 18(6), 925–940.

KOREISHA, S. G., AND T. M. PUKKILA (1989) : "Fast Linear Estimation Methods for Vector Autoregressive Moving-Average Models," *Journal of Time Series Analysis*, 10(4), 325–339.

——— (1990a) : "A Generalized Least-Squares Approach for Estimation of Autoregressive-Moving-Average Models," *Journal of Time Series Analysis*, 11(2), 139–151.

——— (1990b) : "Linear Methods for Estimating ARMA and Regression Models with Serial Correlation," *Communications in Statistics, Part B - Simulation and Computation*, 19(1), 71–102.

——— (1995) : "A Comparison Between Different Order-Determination Criteria for

Identification of ARIMA Models," *Journal of Business and Economic Statistics*, 13(1), 127–131.

KREISS, J.-P. (1985) : "A Note on M-Estimation in Stationary ARMA Processes," *Statistics and Decisions*, 3, 317–336, Correction 4 (1986), 99.

——— (1987) : "On Adaptive Estimation in Stationary ARMA Processes," *The Annals of Statistics*, 15(1), 112–133.

KUROZUMI, E., AND T. YAMAMOTO (2000) : "Modified Lag Augmented Vector Autoregressions," *Econometric Reviews*, 19, 207–231.

LASTRAPES, W. D., AND W. D. MCMILLIN (2004) : "Cross-Country Variation in the Liquidity Effect : The Role of Financial Markets," *Economic Journal*, 114(498), 890–915.

LASTRAPES, W. D., AND G. SELGIN (1995) : "The Liquidity Effect : Identifying Short-Run Interest Rate Dynamics Using Long-Run Restrictions," *Journal of Macroeconomics*, 17(3), 387–404, Summer.

LEHMANN, E. L. (1986) : *Testing Statistical Hypotheses.* John Wiley & Sons, New York, 2nd edn.

LEWIS, R., AND G. C. REINSEL (1985) : "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, 16, 393–411.

LI, H., AND G. S. MADDALA (1996) : "Bootstrapping Time Series Models," *Econometric Reviews*, 15, 115–158.

LÜTKEPOHL, H. (1987) : *Forecasting Aggregated Vector ARMA Processes.* Springer-Verlag, Berlin.

LÜTKEPOHL, H. (1991) : *Introduction to Multiple Time Series Analysis.* Springer-Verlag, Berlin.

——— (1993) : *Introduction to Multiple Time Series Analysis.* Springer-Verlag, Berlin, second edn.

LÜTKEPOHL, H. (1999) : "Forecasting Cointegrated VARMA Processes," Discussion Paper 1, Institut für Statistik und Ökonometrie Humboldt-Universität zu Berlin,

Spandauer Str.1, 10178 Berlin, GERMANY, Tel. : (+49-30) 2093-5718, Fax. : (+49-30) 2093-5712, email : [redacted]

LÜTKEPOHL, H. (2001) : "Vector Autoregressions," in Baltagi (2001), chap. 32, pp. 678–699.

LÜTKEPOHL, H. (2005) : *New Introduction to Multiple Time Series Analysis.* Springer, Springer-Verlag, Berlin Heidelberg, Germany.

LÜTKEPOHL, H., AND H. CLAESSEN (1997) : "Analysis of Cointegrated VARMA Processes," *Journal of Econometrics*, 80(2), 223–239.

LÜTKEPOHL, H., AND D. S. POSKITT (1996) : "Specification of Echelon-Form VARMA Models," *Journal of Business and Economic Statistics*, 14(1), 69–79.

MADDALA, G. S., AND I.-M. KIM (1998) : *Unit Roots, Cointegration and Structural Change.* Cambridge University Press, Cambridge, U.K.

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

MAURICIO, J. A. (2002) : "An Algorithm for the Exact Likelihood of a Stationary Vector Autoregressive-Moving Average Model," *Journal of Time Series Analysis*, 23(4), 473–486.

MCALEER, M., AND L. OXLEY (eds.) (1999) : *Practical Issues in Cointegration Analysis.* Blackwell, Oxford, U.K.

MCMILLIN, W. D. (2001) : "The Effect of Monetary Policy Shocks : Comparing Contemporaneous versus Long-Run Identifying Restrictions," *Southern Economic Journal*, 67(3), 618–636.

MÉLARD, G., R. ROY, AND A. SAIDI (2002) : "Exact Maximum Likelihood Estimation of Structured or Unit Roots Multivariate Time Series Models," Discussion paper, Institut de Statistique, Université Libre de Bruxelles, and Départment de mathématiques et statistique, Université de Montréal.

NEWBOLD, P. (1974) : "The Exact Likelihood Function for a Mixed Autoregressive-Moving Average Process," *Biometrika*, 61, 423–426.

NICHOLLS, D. F., AND A. D. HALL (1979) : "The Exact Maximum Likelihood Function of Multivariate Autoregressive Moving Average Models," *Biometrika*, 66, 259–264.

——— (1980) : "The Evaluation of Exact Maximum Likelihood Estimates for VARMA Models," *Journal of Statistical Computation and Simulation*, 3, 251–262.

NSIRI, S., AND R. ROY (1992) : "On the Identification of ARMA Echelon-Form Models," *Canadian Journal of Statistics*, 20(4), 369–386.

NSIRI, S., AND R. ROY (1996) : "Identification of Refined ARMA Echelon Form Models for Multivariate Time Series," *Journal of Multivariate Analysis*, 56, 207–231.

PAGAN, A. R., AND J. C. ROBERTSON (1995) : "Resolving the Liquidity Effect," *Federal Reserve Bank of St. Louis Review,*, 77, 33–54.

——— (1998) : "Structural Models of the Liquidity Effect," *Review of Economics and Statistics*, 80, 202–217.

PAPARODITIS, E. (1996) : "Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes," *Journal of Multivariate Analysis*, 57, 277–296.

PAPARODITIS, E., AND B. STREITBERG (1991) : "Order Identification Statistics in Stationary Autoregressive Moving-Average Models : Vector Autocorrelations and the Bootstrap," *Journal of Time Series Analysis*, 13(5), 415–434.

PARK, J. Y., AND P. C. B. PHILLIPS (1989) : "Statistical Inference in Regressions with Integrated Processes : Part 2," *Econometric Theory*, 5, 95–131.

POSKITT, D. S. (1987) : "A Modified Hannan-Rissanen Strategy for Mixed Autoregressive-Moving Average Oder Determination," *Biometrika*, 74(4), 781–790.

——— (1992) : "Identification of Echelon Canonical Forms for Vector Linear Processes Using Least Squares," *The Annals of Statistics*, 20(1), 195–215.

POSKITT, D. S., AND H. LÜTKEPOHL (1995) : "Consistent Specification of Cointegrated Autoregressive Moving-Average Systems," Discussion Paper 54, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.

PÖTSCHER, B. (2002) : "Lower Risk Bounds and Properties of Confidence Sets for Ill-Posed Estimation Problems with Applications to Spectral Density and Persistence Estimation, Unit Roots and Estimation of Long Memory Parameters," *Econometrica*, 70(3), 1035–1065.

PUKKILA, T., S. KOREISHA, AND A. KALLINEN (1990) : "The Identification of ARMA Models," *Biometrika*, 77(3), 537–548.

QUINN, B. G. (1980) : "Order Determination for a Multivariate Autoregression," *Journal of the Royal Statistical Society, Series B*, 42, 182–185.

REINSEL, G. C. (1993) : *Multivariate Time Series*. Springer-Verlag, New York.

——— (1997) : *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York, second edn.

REINSEL, G. C., S. BASU, AND S. F. YAP (1992) : "Maximum Likelihood Estimators in the Multivariate Autoregressive Moving-Average Model from a Generalized Least Squares Viewpoint," *Journal of Time Series Analysis*, 13(2), 133–145.

RISSANEN, J. (1978) : "Modelling by Shortest Data Description," *Automatica*, 14, 467–471.

SCHWARZ, G. (1978) : "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

SCOTT, D. J. (1973) : "Central Limit Theorems for Martingales and for Processes with Stationary Increments Using a SKorokhod Representation Approach," *Advances in Applied Probability*, 5, 119–137.

SHEA, B. L. (1989) : "The Exact Likelihood of a Vector Autoregressive Moving Average Model," *Journal of the Royal Statistical Society Series C, Applied Statistics*, 38(1), 161–184.

SIMS, C. (1971a) : "Distributed Lag Estimation When the Parameter Space is Explicitly Infinite-Dimensional," *Annals of Mathematical Statistics*, 42, 1622–1636.

——— (1980) : "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

SIMS, C. A. (1971b) : "Discrete Approximations to Continuous Time Distributed Lags in Econometrics," *Econometrica*, 39, 545–563.

SIMS, C. A., J. H. STOCK, AND M. W. WATSON (1990) : "Inference in Linear Time Series Models with some Unit Roots," *Econometrica*, 58, 113–144.

SOLO, V. (1984) : "The Exact Likelihood for a Multivariate ARMA Model," *Journal of Multivariate Analysis*, 15, 164–173.

——— (1986) : *Topics in Advanced Time Series Analysis*, Lecture Notes in Mathematics. Springer Verlag, New York.

SRIRAM, T. N. (1994) : "Invalidity of Bootstrap for Critical Branching Processes with Immigration," *The Annals of Statistics*, 22, 1013–1023.

STRONGIN, S. (1995) : "The Identification of Monetary Policy Disturbances Explaining the Liquidity Puzzle," *Journal of Monetary Economics*, 35(3), 463–497.

TANAKA, K. (1996) : *Time Series Analysis : Nonstationary and Noninvertible Distribution Theory.* John Wiley & Sons, New York.

TIAO, G. C., AND G. E. P. BOX (1981) : "Modeling Multiple Time Series with Applications," *Journal of the American Statistical Association*, 76(376), 802–816.

TIAO, G. C., AND R. S. TSAY (1985) : "A Canonical Correlation Approach to Modeling Multivariate Time Series," in *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, pp. 112–120, Washington, D.C.

TIAO, G. C., AND R. S. TSAY (1989) : "Model Specification in Multivariate Time Series," *Journal of the Royal Statistical Society, Series B*, 51(2), 157–213.

TODA, H. Y., AND P. C. B. PHILLIPS (1993) : "Vector Autoregressions and Causality," *Econometrica*, 61, 1367–1393.

TODA, H. Y., AND P. C. B. PHILLIPS (1994) : "Vector Autoregression and Causality : A Theoretical Overview and Simulation Study," *Econometric Reviews*, 13, 259–285.

TODA, H. Y., AND T. YAMAMOTO (1995) : "Statistical Inference in Vector Autoregressions with Possibly Integrated Processes," *Journal of Econometrics*, 66, 225–250.

TSAY, R. S. (1989a) : "Identifying Multivariate Time Series Models," *Journal of Time Series Analysis*, 10(4), 357–372.

TSAY, R. S. (1989b) : "Parsimonious Parameterization of Vector Autoregressive Moving Average Models," *Journal of Business and Economic Statistics*, 7(3), 327–341.

TSAY, R. S. (1991) : "Two Canonical Forms for Vector ARMA Processes," *Statistica Sinica*, 1, 247–269.

TSAY, R. S., AND G. C. TIAO (1985) : "Use of Canonical Analysis in Time Series Model Identification," *Biometrika*, 72(2), 299–315.

WIENER, N. (1956) : "The Theory of Prediction," in *The Theory of Prediction*, ed. by E. F. Beckenback, chap. 8. McGraw-Hill, New York.

YAMADA, H., AND H. Y. TODA (1998) : "Inference in Possibly Integrated Vector Autoregressive Models : Some Finite Sample Evidence," *Journal of Econometrics*, 86, 55–95.

YAMAMOTO, T. (1996) : "A Simple Approach to the Statistical Inference in Linear Time Series Models Which May Have Some Unit Roots," *Hitotsubashi Journal of Economics*, 37, 87–100.

ZHAO-GUO, C. (1985) : "The Asymptotic Efficiency of a Linear Procedure of Estimation for ARMA Models," *Journal of Time Series Analysis*, 6(1), 53–62.

Conclusion générale

Dans cette thèse nous avons développé de nouvelles méthodes paramétriques permettant d'établir de l'inférence fiable à distance finie dans les modèles VAR. Ces méthodes sont basées sur des techniques de simulation utilisant les tests de MC. Nous avons également développé de nouvelles techniques simples pour la spécification et l'estimation des modèles VARMA sous la forme échelon. Cette dernière étant attrayante de point de vue théorique de part sa représentation parcimonieuse accusait un refus de part les économètres sur le plan pratique [difficile à utiliser, en particulier dans de grands systèmes]. Les deux derniers essais de cette thèse offrent des nouveaux outils rendant possible l'estimation et l'inférence (basée sur les nouvelles méthodes développées déjà pour les VAR) dans les modèles VARMA en forme échelon. Cette thèse est composée de trois essais.

Dans le premier essai, nous avons proposé une méthode basée sur la simulation pour produire des tests à échantillons finis dans les modèles VAR paramétriques avec ordre connu (ou avec borne supérieure connue de l'ordre du processus). La caractéristique importante de cette méthode est qu'aucune autre hypothèse sur la structure du processus sous-jacent n'est nécessaire : tout ce qui est nécessaire, c'est la possibilité de simuler le modèle une fois qu'un nombre fini de paramètres a été spécifié. Par exemple, le processus VAR peut être intégré de n'importe quel ordre. Nous avons également montré que la méthode proposée peut être appliquée dans la pratique, malgré la présence d'un grand nombre de paramètres de nuisance. À l'aide d'une étude de simulation, nous avons vu clairement que les deux procédures asymptotique standard ainsi que le bootstrap peuvent souffrir de graves distorsions de niveau, tandis que, sous les mêmes conditions, la méthode de MCM permet de contrôler parfaitement le niveau du test (comme prévu), bien que la taille pourrait être inférieure au niveau du test. À notre connaissance, aucune autre procédure disponible, ne possède ces caractéristiques. Nous avons aussi appliqué la méthode pour des tests de causalité au sens de Granger dans un modèle VAR comprenant quatre variables macroéconomiques de l'économie américaine.

Bien que nous avons considéré ici des tests sur l'ordre du VAR et de causalité au sens de Granger, l'approche proposée peut être appliquée en principe à tout ensemble de restrictions sur le modèle, telles que des hypothèses de racine unitaire ou de cointégration. Dans de tels cas, même si l'hypothèse de racine unitaire (par exemple) pourrait être prise en compte par une théorie distributionnelle asymptotique ou d'une procédure de bootstrap, de grandes racines dans la région stationnaire mais proches du cercle unité pourraient, encore, conduire à de grandes distorsions de niveau. Par construction, la procédure MCM reste valable indépendamment de la structure du VAR. Il est également important de noter que la distribution des innovations n'a pas besoin d'être normale : toute hypothèse qui spécifie complètement la distribution de $\varepsilon(T) = \text{vec}(\varepsilon_1, \ldots, \varepsilon_T)$, *i.e.* c'est-à-dire la distribution des perturbations à une transformation linéaire (ou matrice de covariance) près inconnue, peut être utilisée. Aucune hypothèse sur l'existence des moments n'est nécessaire, donc on pourrait envisager des distributions avec des queues épaisses. On pourrait aussi introduire d'autres paramètres libres dans la distribution des erreurs : ces paramètres peuvent être traités comme des paramètres de nuisance supplémentaires.

Les principales limites de l'approche que nous avons proposée ici se situent dans la conception paramétrique requise pour effectuer les tests de MC et le coût de calcul. Sur le premier point, il est important de noter que les hypothèses paramétriques suppose une borne maximale sur l'ordre du processus (qui est équivalent à supposer que l'ordre du processus VAR est "connu"). Dans le cas des tests de non causalité au sens de Granger (ainsi que pour de nombreuses hypothèses d'intérêt), cela signifie que l'ordre du retard est une partie intégrante de l'hypothèse nulle : il n'y a aucun moyen de "séparer" la non causalité au sens de Granger de l'ordre du processus. En prévoyant une sélection de l'ordre fondée sur les données cela exigerait la simulation de la procédure de sélection de modèle. Notez, toutefois, que faire de l'inférence à échantillon fini sans mettre explicitement une borne supérieure sur l'ordre du processus est une tâche fondamentalement impossible [voir les débats dans Sims (1971a, 1971b), Cochrane (1991), Blough (1992), Faust (1996, 1999), Pötscher (2002) et Dufour (2003)]. Donc, du point de vue de l'élaboration de tests valides à échantillons finis, l'hypothèse d'un

"ordre connu" est inévitable.

Si l'on est prêt à accepter une procédure qui ne possède pas une "justification asymptotique", il est également important de noter que les "procédures exactes" proposées restent asymptotiquement valides (dans le sens habituel de validité asymptotique ponctuelle) sous des hypothèses beaucoup plus faibles, y compris un ordre "inconnu" qui peut être "estimé d'une manière convergente". Tant que les tests MC sont effectués au moyen d'une distribution qui est couverte par les hypothèses de la théorie distributionnelle limite, la probabilité de l'erreur du type I satisfait la condition de niveau, asymptotiquement. Bien évidemment, sous des hypothèses habituelles, telle que la convergence ne va pas généralement être uniforme – ce qui suppose la possibilité de déviations arbitraires par rapport au niveau nominal du test – mais cela reflète simplement le fait que les hypothèses générales de régularité sont tout simplement trop faibles pour permettre même l'existence de procédures exactes valides à échantillons finis [voir Dufour (2003)]. Il est également intéressant de noter que la procédure MCM s'adapte automatiquement à la dépendance éventuelle de la distribution de la statistique du test sur les coefficients autorégressifs.

Concernant le deuxième point, il est clair que les tests MCM tests sont intensifs côté calcul. Le code qui nous permet de réaliser des simulations et des applications présentées n'est pas certainement optimal [étant donné que ceux-ci étaient réalisés avec GAUSS] et nous nous efforçons de l'améliorer. Compte tenu de l'amélioration régulière de la vitesse des ordinateurs des générations futures, l'importance de cette limitation devrait diminuer dans le temps.

Dans le deuxième essai, nous avons donné la distribution asymptotique d'un estimateur linéaire simple de deux étapes pour les modèles VARMA en forme échelon. L'estimateur est convergent lorsque la longue autorégression auxiliaire utilisée pour générer les estimations de première étape des innovations du modèle, a un ordre n_T qui croit vers l'infini à un taux inférieur à T^{δ} où $0 < \delta_0 \leq \delta < 1/2$. De plus, cet estimateur a une distribution asymptotique normale lorsque n_T croit à un taux inférieur à T^{δ} avec $0 < \delta_0 \leq \delta < 1/4$. Dans ce dernier cas, la distribution asymptotique n'est pas affectée par l'utilisation des résidus estimés décalés comme régresseurs.

Les résultats ci-dessus peuvent être exploitées de plusieurs manières. Tout d'abord, les estimations de deux étapes et la théorie distributionnelle correspondante peuvent être utilisées directement pour l'inférence sur le modèle VARMA. En particulier, ces résultats peuvent être utilisés pour fin de sélection de modèle et de simplification du modèle (par exemple, en supprimant les coefficients non significatifs). Deuxièmement, les estimations de deux étapes peuvent être exploitées pour obtenir des estimateurs plus efficaces, tels que les estimateurs MV ou des estimateurs qui sont asymptotiquement équivalents au MV. Cela peut être considéré, en particulier, afin d'atteindre l'efficacité avec des innovations Gaussiennes. Notons, toutefois, que de tels gains d'efficacité ne peuvent être obtenus si les innovations ne sont pas Gaussiennes. Troisièmement, en raison de leurs simplicité, les estimateurs linéaires de deux étapes sont particulièrement bien adaptés pour être utilisés dans le contexte de procédures d'inférence basées sur les techniques de simulation, comme les tests de bootstrap et MCM. En outre, la distribution asymptotique donnée ci-dessus peut être utile afin d'améliorer la validité du bootstrap.

Dans le troisième et dernier essai de cette thèse, nous avons proposé une procédure d'estimation linéaire de trois étapes pour l'estimation des modèles VARMA stationnaires inversibles, en forme échelon. L'approche peut être facilement adaptée aux modèles VARMAX et étendue aux modèles VARMA intégrés et cointégrés. La méthode d'estimation a considéré la forme échelon car celle-ci tend à identifier des modèles ayant une paramétrisation relativement parcimonieuse. Mais notre procédure reste valable pour d'autres formes identification, telle que la forme d'équations finales ou tout autre modèle contraint pour fins d'inférence.

Notre méthode proposée fournit une forme standard générale simplifiée des estimations des paramètres de la forme échelon qui sont plus faciles à calculer que celles de Hannan et Kavalieris (1984b) et Poskitt (1992), respectivement. Ce qui parait recommandable et attrayant du point de vue pratique. En outre, nous avons généralisé les résultats développés dans le deuxième essai sur la méthode d'estimation de deux étapes pour dériver la distribution asymptotique des estimateurs MCG en cas de bruits blancs forts, car cela n'a été nulle part établi à notre connaissance. Nous

avons également fourni la justification théorique de la mise en œuvre de la troisième étape de régression linéaire pour permettre l'efficacité en cas d'erreurs Gaussiennes. Ce qui montre l'équivalence entre la distribution asymptotique de nos estimateurs de troisième étape et celle des estimateurs MV. Cependant, les propriétés à échantillon fini de nos estimateurs ne sont pas les mêmes que celles des estimateurs MV. Bien que notre procédure d'estimation de trois étapes est équivalente à celle de Hannan et Kavalieris (1984b), les estimations des covariances asymptotiques des paramètres de la forme échelon que nous avons donnés pour les deuxième et troisième étapes de la procédure, sont par contre simples, agréables et faciles à utiliser pour fins d'inférence.

Dans cet essai, nous avons également proposé une procédure simplifiée de sélection d'ordre (ou de modèle) pour identifier les indices de Kronecker caractérisant la forme échelon du modèle VARMA. Notre procédure repose sur la détermination de la matrice implicite des restrictions pour tous les ensembles d'indices de Kronecker et pour n'importe quelle dimension possible des systèmes VARMA. Ainsi, par un tel accomplissement nous avons rendu la forme échelon plus attrayante pour des travaux empiriques futurs. En outre, pour résoudre les problèmes de sur-paramétrisation et de précision dont souffrent, respectivement, les méthodes de Hannan et Kavalieris (1984b) et de Poskitt (1992), nous avons proposé des critères d'information et des raccourcis pour fournir des estimations fortement convergentes des vraies valeurs des indices de Kronecker dans des échantillons de taille modérée et grande. Par ailleurs, afin d'estimer des modèles VARMA en forme échelon, stationnaires et inversibles (moyennant entre autres notre procédure d'estimation), nous avons développé un algorithme efficace qui fonctionne d'une manière systématique à cette fin. Un tel algorithme pourrait être facilement intégré pour améliorer la librairie (routines ou Logiciels) de maximisation du MV en vue d'estimer des modèles VARMA stationnaires et inversibles. De plus, une étude de simulation a montré que notre méthode d'estimation domine de loin celles de Hannan et Kavalieris (1984b) et de Poskitt (1992), respectivement. Il en est ainsi, en particulier, et davantage dans des systèmes avec des dimensions élevées. C'est, à dire, qu'à la différence de leurs approches qui se comportent mal dans l'estimation des indices de Kronecker, présentant ainsi de

mauvaises propriétés à échantillon fini des estimations des paramètres impliqués par la forme échelon, notre procédure, en revanche, donne des probabilités élevées pour identifier le vrai modèle VARMA sous la forme échelon: En outre, notre procédure fournit des paramètres estimés avec des propriétés à échantillon fini agréables et assez satisfaisantes. Donc, au vu des résultats ci-dessus, notre méthode d'estimation proposée semble plus précise et plus recommandable dans la pratique.

Enfin, à l'aide d'un exemple empirique sur l'économie américaine utilisant six séries macro-économiques, nous avons facilement estimé un modèle VARMA en forme échelon avec la méthode que nous avons proposée. De plus, nous avons montré que ces modèles plus parcimonieux fournissent de meilleurs outils efficaces, comparés aux modèles VAR, dans l'analyse de certaines politiques économiques telle que de la politique monétaire que nous avons étudiée.