SEASONAL COINTEGRATION IN MACROECONOMIC SYSTEMS: CASE STUDIES FOR SMALL AND LARGE EUROPEAN COUNTRIES

Robert M. KUNST

Forschungsbericht/
Research Memorandum No. 271

Oktober 1990

ABSTRACT

Six-variable vector autoregressive systems consisting of macroeconomic series are investigated. Parallel data series for four European countries are used: Austria, Germany (Federal Republic), Finland, and the United Kingdom. All data series are not seasonally adjusted.

The aim of the paper is to show that most of the series are better modeled using stochastic seasonality and seasonal unit roots models than simple deterministic models of seasonal structures. As a second step, seasonal cointegration in the systems is studied. It is shown that all four economies display seasonal cointegration as well as usual cointegration.

ZUSAMMENFASSUNG

Gegenstand der Untersuchung sind sechs-dimensionale vektorautoregressive Systeme. Hiezu werden Paralleldaten aus vier europäischen Staaten herangezogen: Österreich, Deutschland (Bundesrepublik), Finnland, und das Vereinigte Königreich. Die Daten liegen in nicht-saisonbereinigter Form vor.

Ziele des Papiers ist es zu zeigen, daß die meisten Zeitreihen besser durch stochastische Saison und saisonale Einheitswurzeln als durch einfache deterministische Saisonmodelle beschrieben werden. Als zweiter Schritt wird saisonale Kointegration in den Systemen studiert. Es wird gezeigt, daß alle vier Ökonomien sowohl saisonale als auch die übliche (Frequenz 0) Kointegration aufweisen.

1. Introduction

For some time, vector autoregressions (VAR) have been in use now as the basic mechanical method for description and prediction of multivariate economic time series. A rather lengthy dispute between those who advocate the use of differenced data and those who prefer modeling original ("level") series has been all but reconciled by the theory of cointegration [Engle and Granger (1987)]. Until recently, detailed and readily applicable results were available for first-order integrated systems only but extensions for higher-order systems and the seasonal case are in the focus of current research activities [e.g. see Engle, Granger, and Hallman (1989), Hylleberg, Engle, Granger, and Yoo (1990), or Sims, Stock, and Watson (1990)].

For the moment, let us start from the first-order integrated vector autoregressive system which is defined as a vector autoregression (VAR) with roots outside the unit disc or at one only. It is assumed explicitly that all individual series can be made stationary by first-order differences. A generic feature of these models is the phenomenon of first-order cointegration (CI(1,1) in the notation of Engle and Granger (1987)). The solution of the Gaussian maximum-likelihood problem for this case has been analyzed by Johansen (1988) and Johansen and Juselius (JJ, 1989).

Raw (non-adjusted) monthly or quarterly economic time series frequently show seasonal patterns which shed some doubt on the assumption of stationary first differences. The question whether these seasonal patterns should be eliminated by regression on seasonal dummies (the "deterministic" model) or by treating them by seasonal differencing, thereby assuming additional unit roots on the unit circle (the "stochastic" model), parallels the discussion of deterministic and stochastic trend models. If the stochastic model is accepted, again cointegration arises as a generic feature, this time cointegration at seasonal frequencies.

Cointegration at seasonal frequencies means that, although individual series display stochastic seasonality giving rise to unit roots e.g. at -1 (frequency π), there is a linear combination which is free from that kind of seasonality but might yet have unit roots at 1 or at other seasonal frequencies (e.g. $\pi/2$ in this example). Up to now, evidence on seasonal cointegration has been rather scarce (compare the recent contribution by Engle, Granger, Hylleberg, Lee (1990)). The examples of this paper show that seasonal cointegration is not uncommon in larger systems whereas the author must concede that it will be difficult to encounter in two-variable relations.

The object of this paper's investigations are six-dimensional macroeconomic data sets from four European economies (Austria, Federal Republic of Germany, Finland, United Kingdom). The six series are: gross domestic (or national) product; private consumption; gross fixed investment; goods (if unavailable, total) exports; real interest rate on bonds; real wage. This system coincides with the one used by Kunst and Neusser (1990) for the Austrian economy. They motivate their specification by the fact that it contains four cointegrating relationships, according to neoclassical growth theory (real interest, consumption to output quota, investment to output quota, real wage to output ratio), with exports additionally contributing as an important factor in modern open economies. Therefore, it serves as an appropriate starting point for investigating multivariate cointegration in many countries although it may be conceded that, in some economies, other variables could play a more important role than some of those included.

The paper is organized as follows. The subsections of Section 2 are concerned with the univariate properties of the 24 data series. (Sub)Section 2.1 discusses the property of integration at frequency zero, deterministic versus stochastic trend models. Section 2.2 expounds a parallel discussion on stochastic versus deterministic seasonality and quotes the relevant test statistics. Section 2.3 allows for a short look at the distributional properties of the data series which can be thought of as tests for outliers. Section 2.4 describes the data. Section 2.5 gives the results of the seasonal unit root tests and shows that seasonal unit roots are not uncommon among economic series.

Section 3.1 quotes the maximum likelihood estimation algorithm of Johansen (1988) for cointegrated vector autoregressions whereas Section 3.2 quotes the extensions by Lee (1989) for the seasonally cointegrated case. Section 3.3 takes a closer look at the results from the seasonal cointegration testing and estimation procedure. Section 3.4 gives the frequency zero cointegration vectors which emerge as a byproduct from the analysis. A companion paper focuses on these vectors in an inter-country comparison [Kunst (1990)]. Section 3.5 gives a short summary of an experiment considering seasonal cointegration and deterministic dummy structures at the same time. Section 4 concludes.

2. Univariate characteristics of the data series

2.1 Integration at frequency zero

For a long time, applied economics used trend-stationary models to represent typical macro-economic series, such as output or consumer expenditures. The trend-stationary model views the data at hand as the sum of a deterministic long-run growth component and a stationary (stochastic) business cycle component, the latter one being the only part which is of genuine economic interest. The obvious remedy is to regress the raw series on some basic deterministic shapes, preferably linear and maybe quadratic time trends, and focusing univariate as well as multivariate analysis on the residual. Such models may look like the following one: [1]

$$\begin{aligned} X_t &= a + bt + C_t \\ \Phi(B)C_t &= \theta(B)\epsilon_t \end{aligned} \qquad (2.1)$$

In (2.1), the roots of the (determinants of the) lag polynomials $\Phi(.)$ and $\theta(.)$ are restricted to lie outside of the unit disk and $\{\epsilon_t\}$ is white noise (serially uncorrelated). In most cases, $\{X_t\}$ will be the log of the original data and (2.1) thus implies that the original series fluctuates around an exponential trend.

In recent years, the prevalence of models like (2.1) in empirical economics has waned and unit root models, also called integrated

1 B will denote the backshift or lag operator in the following.

or difference-stationary models, have gained widespread popularity. The typical model of this type looks

$$\Phi(B)\Delta X_t = b + \theta(B)\epsilon_t \qquad (2.2)$$

In (2.2), $\Phi(.)$, $\theta(.)$, and $\{\epsilon_t\}$ follow the same restrictions as in (2.1) and Δ is used in short for 1-B. Eliminating Δ from the equation, i.e. integrating it, leads to a similar deterministic component a+bt as in (2.1) but now the deviations from this "trend" are non-stationary.

As the question of whether deviations from the trend are stationary or not has profound economic implications, e.g. on the "persistence" of unanticipated shocks, the discrimination problem among (2.1) and (2.2) has led to many different testing procedures and much empirical work on economic series from many countries. In general, evidence seems to prefer (2.2) but this is still not being agreed upon universally, even with respect to series investigated very thoroughly such as U.S. real GNP. It seems that, given the finite samples available, the question cannot be decided empirically.

In the following, all series will be *assumed* to be unit root processes. Evidence favors this working hypothesis in most cases as is seen from the first column of the statistics in Table 2. It may be discomforting for some economists to see the effects of shocks persist forever in the series but, on the other hand, it makes more sense to have mean-reverting growth rates than to believe in a mystic long-run trend to which the processes are attracted by some magical principle. Interestingly, forecasting practitioners tend to use second-order integrated models implicitly by adjusting medium-run growth scenarios to the last available information [2]. Except for some wage and price series, integrated growth rates are, however, safely rejected by virtually all statistical procedures.

A last argument for the use of integrated models is of a more game

2 In years of recession, output growth predictions for following years are usually tracked down to 1-2 % while this figure rises to more than 3 % in years of high economic activity. The trend-stationary model would demand for exactly the opposite reaction.

theoretic nature. In a multivariate system, even one integrated series suffices for the statistical procedures expounded in Section 3 to be useful whereas the (stationary) modeling of residuals from regressing integrated series on deterministics invalidates all results thoroughly.

2.2 Seasonality

The reason for the discussion on trend-stationary versus integrated models in the last section to be a bit more lengthy than seems to be fit for this paper has been to enable drawing parallels between the treatment of trends and of seasonality. Again, the usual conception has been that raw measurements of monthly or quarterly series can be decomposed additively - the frequent use of logged data implies a multiplicative components model on the original series - into a possibly deterministic seasonal component and into the economically interesting remainder, this in turn maybe consisting of "trend" and "cycles".

$$X_t = S_t + E_t \tag{2.3}$$

Here, the second component is labeled E_t because it is supposed to be economically interesting. Any univariate and multivariate modeling is usually performed on these $\{E_t\}$ processes and it is assumed that these contain all the relevant information of $\{X_t\}$.

A principal difference to the trend problem is that components modeling has actually gone much farther in the sense that, in many countries for many variables, the $\{X_t\}$ series is not even available. All seasonal information is destroyed carefully by specialists, often on a low level of aggregation (e.g. for agricultural exports and not for the gross national product). Recent years have seen, however, a general increase of interest in the raw series. Within Europe, the criterion of the availability of unadjusted series has confined this investigation to four countries: Austria, Finland, Germany (Federal Republic), and the United Kingdom.

The decomposition in (2.3) is the backbone of the so-called seasonal adjustment procedures, all of them designed to clean the series at hand from seasonal noise. The most well-known and internationally standardized procedure is Census X-11 which is used by many statistical offices in various versions. The use of these non-linear filters brings us back into a non-seasonal E_t-world. Once original data are available, many econometricians prefer to use simple deterministics, such as quarterly [3] dummies or dummy trends, in accordance with linear time series analysis on the stochastic part of the process or system. The specification with seasonal dummies

$$X_t = \sum_{i=1}^{4} a_i D_{it} + \Phi^{-1}(B)\theta(B)\epsilon_t \quad (2.4)$$

is the exact replica of the deterministic trend modeling in (2.1).

The counterpart of the integrated modeling of (2.2) in the presence of seasonal effects is the application of seasonal differences

$$\Phi(L)\Delta_4 X_t = \theta(L)\epsilon_t \quad (2.5)$$

Here, Δ_4 is used to denote $1-B^4$ which has a direct interpretation if $\{X_t\}$ is a logged series as $\Delta_4 X_t$ then is the annual growth rate. Δ_4 mathematically consists of the two factors $1-B$ and $1+B+B^2+B^3$, the former one of which removes the trend while the latter one removes all seasonal structure. An application of Δ_4 together with an extra Δ is rarely justified. Model (2.5) is a member of the SARIMA (seasonal integrated ARMA) class defined by Box and Jenkins (1976). In the following, the more precise terminology suggested by HEGY is adopted here and the process $\{X_t\}$ is rather called "integrated at the frequencies 0, $\pi/2$, π", as these are the frequencies of the spectral poles implied by the unit roots at +1, $\pm i$, -1 respectively, these in turn being the roots of the Δ_4 operator.

It is important to understand the difference between the two seasonal models. (2.4) describes the behavior of a series that is governed by four alternating deterministic trends with identical

3 All formulae in this paper assume that seasonal data are quarterly. In case of monthly series, some of the analysis becomes slightly more cumbersome but remains similar in principle.

slopes, i.e. parallel lines. This means that the original shape of the seasonal pattern will remain constant. (2.5), on the other hand, in its extreme form with uncorrelated errors $\Phi(.)\equiv\theta(.)\equiv1$ describes four circularly merged random walks with the same drift constant, which implies persistent changes in the seasonal pattern although the best prediction of its future shape will always be its present one. Neither of the two models accounts for monotonously expanding seasonal patterns though a model using dummy trends instead of dummy constants would. Visual evidence prefers the stochastic model in some cases but the seasonal pattern of many other series does not appear to change enough to reject the deterministic model (2.4). Compare the Austrian wage series in Figure 6 for a good example for a changing seasonal pattern.

At this point, anyone unfamiliar with the problem is likely to predict that the research paradigm with respect to seasonality will be subject to the same development as with respect to detrending: discriminatory tests will be developed which corroborate the stochastic model and the deterministic model finally will be all but abandoned in favor of the seasonal unit roots model. Interestingly, the first prediction holds true while the second one does not, this being due to several reasons.

Firstly, the pure model $\Phi(.)\equiv\theta(.)\equiv1$ (the "seasonal random walk") is not very attractive. If started from a flat seasonal pattern, it tends to produce too many changes in seasonality, summer peaks becoming winter peaks and spring troughs turning into autumn troughs too easily. Against this, it can be argued that $\Phi(.)\equiv\theta(.)\equiv1$ is unrealistic for economic series and that *annual* growth rates usually exhibit substantial positive correlations among *quarters*. If additionally a distinct seasonal starting pattern of high volatility relative to the innovations sequence is used, summer peaks turning into winter peaks will be a rare sight in simulated series. One should not be too reluctant, however, to attach positive probability e.g. to the event of shifting the main feast from Christmas to summer solstice in Europe after a time span of one or two centuries.

Secondly, seasonal data are still not accessible in many cases, as some series is not due to isolated outliers but rather to true leptokurtic volatility. Some of the statistics could be affected by this phenomenon.

TABLE 1: Skewness and kurtosis of seasonal differences of investigated (logged) series

	Austria		Finland	
	Skewness	Kurtosis	Skewness	Kurtosis
Y	-.016	-.508	-.213	.232
C	-.267	1.082*	.024	-.635
I	.145	.481	-.204	.598
X	.210	1.002	-.872**	1.331*
R	-.260	.650	.358	2.730**
W	-.016	.112	-.256	-.024
	Germany		**United Kingdom**	
	Skewness	Kurtosis	Skewness	Kurtosis
Y	-.149	-.260	.153	2.006**
C	-.081	-.477	-.057	-.415
I	-.236	1.036*	.088	2.101**
X	-.173	.335	.010	1.164*
R	.271	.597	-.356	2.980**
W	.064	-.441	-.841**	.275

Note: * indicates significance at 5 %, ** at 1 %.

2.4 The data

For each country, data consist of 6 quarterly series on: real gross domestic product (Y), real private consumption (C), real gross investment (I), real interest rate (deflated bond rate R), real goods exports (X), real wages (deflated per capita wages W). Countries investigated are Austria, Finland, Germany (Federal Republic), and the United Kingdom. The selection of variables has been adopted from Kunst and Neusser (1990) who investigate cointegrating structures in the Austrian system. The selection of countries has been implied by the requirement that quarterly national accounts data be available which have not been seasonally adjusted. In many countries (including e.g. the United States), only seasonally adjusted accounts are published. Anyway, the cases include two large and two small European economies and thus represent an interesting selection. Data series start at first quarters of the following years: Austria 1964, Finland 1972, Germany 1960, United Kingdom 1963. All series end at 1987/88 with the end of the availability of data. The four national accounts series Y, C, I, X are in constant prices, so that the whole system is in real terms. All series except R are used in logarithms. All series are displayed graphically in Figures 1 to 6.

2.5 Results of the univariate unit root tests

HEGY tests have been performed for all 24 series in the sample. The results are summarized in Table 2. Two of the many different specifications of deterministics have been selected for the table: firstly, linear trend plus seasonal constants; secondly, linear trend alone. In the first case, all alternative hypotheses are modeled explicitly in the test. Additionally, test statistics from a test on the full seasonal difference factor are given, these based on a regression without any deterministics save an intercept.

The test results show that seasonal unit roots and thus stochastic seasonality are very common phenomena indeed among economic series. Evidence *against* stochastic seasonality is strongest in the United Kingdom where it appears to be restricted to consumer expenditures. On the other hand, all German series display stochastic seasonal cycles and rejection of the Δ_4 operator in the R series is rather due to the absence of the factor Δ. The real interest rate is seasonally infested as it has been deflated via the GDP deflator. The officially published deflator of the United Kingdom is non-seasonal and this is reflected in non-seasonal R. Austria and Finland are intermediate cases, the non-seasonal nature of goods exports is corroborated by inspection of Figure 4.

Interestingly, the insertion of seasonal dummies is of less help in coping with seasonal cycles than would have been expected. The fixed cycles are a good model for the investment series where construction investment is low during the winter due to climatic reasons. They are maybe also adequate for a characterization of British GNP. For the rest of the sample, differences among the two versions of the HEGY test are erratic or absent. In summary, with the stated exceptions, seasonality is either absent or stochastic.

The column headed p in Tables 2a-b indicates the number of conditioning (augmenting) lags used in the HEGY regression. p was fixed by first fitting autoregressive models in levels and then subtracting four from the minimum lag order set by the requirement of clean residuals. The Austrian and some of the German lag orders are disturbingly high which indicates the possibility that these series might be better approximated by moving-average (or mixed) than by autoregressive structures.

Additional to these tests, conventional unit root tests have been performed on the original, on seasonally adjusted, and on differenced series. The overall results are that the assumption of first-order integration at frequency zero is consistent with all series except for some cases which may be stationary already and that no indication is given on higher-order integration which would have been harmful for the cointegration analysis. The detailed results are not given in order to save space.

TABLE 2a: Seasonal unit root test statistics

	t_1	t_2	t_3	t_4	F_{34}	$F_{\blacktriangle 4}$	p
Austria							
Y	-.96	-3.10**	-1.81	-1.31	2.50	3.55**	5
	-1.01	-.16	-.95	-.88	.84		
C	-.80	-2.50	-1.79	-1.63	2.90	2.22	5
	-.95	-.75	-1.33	-.73	1.15		
I	-1.88	-3.58***	-3.52**	-.57	6.44*	1.02	4
	-1.71	-.46	.01	.06	.00		
X	-.19	-3.73***	-4.58***	-3.12***	15.38***	10.68***	5
	-.51	-1.27	-2.72***	-1.89*	5.46***		
R	-2.44	-2.62*	-3.08	1.36	5.86*	5.17***	2
	-2.43	-1.53	-2.70***	1.35	4.72***		
W	-1.05	-3.26**	-1.66	-1.35	2.36	3.01*	6
	-1.03	-.51	-1.95**	-1.68*	3.45**		
Germany							
Y	-1.52	-2.83*	-2.28	-1.15	3.29	1.56	1
	-1.56	-1.06	-1.22	-1.11	1.37		
C	-1.21	-1.45	-2.51	-.65	3.41	1.50	4
	-1.23	-.59	-1.53*	-.15	1.18		
I	-2.41	-2.86*	-4.30***	-.58	9.41***	.92	1
	-2.41	-.38	-1.05	-.19	.57		
X	-.13	-1.66	-2.86	-.92	4.60	2.97*	8
	-.04	-.68	-1.30	-.58	1.03		
R	-3.76**	-1.65	-2.59	-1.20	4.07	3.23*	7
	-3.85**	-.11	-1.24	-.57	.93		
W	-1.18	-1.62	-1.36	-1.55	2.18	2.33	4
	-1.08	.51	.76	-1.04	.82		

Note: First row gives HEGY test result with trend and seasonals included; second row with trend only. t_1, t_2, t_3 and t_4 test for the roots +1,-1,±i; F_{34} tests for ±i; $F_{\blacktriangle 4}$ tests for all four roots jointly. * denotes significance at the 10 %, ** at the 5 %, and *** at the 1 % level.

TABLE 2b: Seasonal unit root test statistics

	t_1	t_2	t_3	t_4	F_{34}	$F_{\blacktriangle 4}$	p
Finland							
Y	-1.97	-1.28	-3.18	-.90	5.46*	0.69	1
	-2.05	1.04	-1.30	.37	.91		
C	-1.39	-2.83*	-3.03	-1.79	6.20*	2.61	1
	-1.51	-.99	-2.38**	-1.39	3.79**		
I	-1.62	-2.04	-4.28**	-.01	9.15**	0.36	1
	-1.50	-.45	-.68	.37	.30		
X	-1.83	-3.71**	-5.68***	-1.27	17.04***	4.15**	1
	-1.99	-1.77*	-3.62***	-.37	6.60***		
R	-2.98	-1.88	-4.93***	1.82	13.82***	1.89	1
	-2.50	-1.84*	-1.52*	.36	1.22		
W	-3.84**	-3.23**	-2.59	1.18	4.05	3.42**	3
	-4.77***	-2.58**	-.89	.90	.80		
U.Kingdom							
Y	-2.24	-2.20	-4.12***	-1.76	10.09***	1.43	1
	-2.27	-1.75*	-1.36	-.66	1.14		
C	-2.61	-1.97	-3.08	-1.05	5.40	.48	5
	-2.67	-.97	-.16	-.74	.29		
I	-2.12	-3.37**	-4.97***	-.91	12.94***	2.56	1
	-1.88	-2.47**	-1.75*	-.00	1.53		
X	-1.33	-2.74*	-3.70**	-2.22*	10.35***	6.35***	4
	-1.24	-1.79*	-3.07***	-2.21**	7.83***		
R	-2.79	-3.16**	-6.70***	.28	22.49***	14.95***	1
	-2.77	-3.20***	-6.32***	.30	19.99***		
W	-1.53	-2.93**	-3.88**	-3.93***	15.18***	9.39***	1
	-1.59	-2.57**	-3.82***	-3.78***	14.39***		

Note: see Table 2a

3. The Maximum Likelihood Estimator for cointegrated VAR systems

3.1 VAR systems integrated at frequency zero

The following solution of the problem of estimating a VAR system with cointegrating restrictions is due to Johansen (1988). The relevant statistical foundations are found in Tso (1981). Only the main results will be reviewed here.

Assume that the VAR system is given in the following form

$$X_t = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \dots + \pi_k X_{t-p} + \epsilon_t \tag{3.1}$$

Then, without any assumptions on its stability, it can be rewritten in differences ($\Delta X_t = X_t - X_{t-1} = (1-B)X_t$) but with lag order reduced by 1 plus a matrix which takes care of the fact that the unit factor 1-B possibly is not a factor of all polynomial elements:

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \dots + \Gamma_{p-1} \Delta X_{t-p+1} + \Gamma_p X_{t-p} + \epsilon_t \tag{3.2}$$

Generally, 1-B will not be a factor at all and Γ_p will be a full-rank matrix. If we restrict attention to systems of first-order integrated variables and assume ΔX_t to be stationary, the rank of Γ_p necessarily is less than the system dimension, say, n. It can be shown that, in this case, Γ_p can be represented in the form $\alpha\beta'$ with the factor matrices α and β being nxr-matrices with full smaller rank r where r = rank Γ_p. This representation is unique except for a transformation by a non-singular rxr-matrix.

β has a straightforward interpretation as its columns contain linearly independent cointegrating vectors. The matrix α distributes the influence of the implied stationary error-correction variables $\beta' X_t$ to the components of ΔX_t.

The solution to the problem of efficiently estimating the parameters in the Γ_i matrices under the restriction rank $\Gamma_p = r$ on the basis of Gaussian white noise errors is obtained by the following steps:

1) Regress ΔX_t on $\Delta X_{t-1},\ldots,\Delta X_{t-p+1}$ (least squares equation by equation)
2) Regress X_{t-p} on $\Delta X_{t-1},\ldots,\Delta X_{t-p+1}$
3) Calculate the canonical correlations between the residuals from steps 1 and 2. The eigenvectors corresponding to the non-zero correlations are the columns of β.
4) An estimate for α is obtained from $S_{0p}\beta$ with S_{0p} being the cross-moments matrix of the residuals from steps 1 and 2.
5) Retrieve estimates for the $\Gamma_1,\ldots,\Gamma_{p-1}$ from regressing $\Delta X_t-\alpha\beta' X_{t-p}$ (using the estimates for α and β) on $\Delta X_{t-1},\ldots,\Delta X_{t-p+1}$.

These estimates can be shown to be the maximum likelihood estimates and to be consistent of different order (β consistent of order T, the remainder of order $T^{\frac{1}{2}}$). The variance matrix of the ϵ_t can also be estimated by the same procedure.

The canonical correlations or roots calculated in step 3 are important. Decisions about whether they are zero or not can be based on the likelihood-ratio (LR) statistic

$$-T\ \Sigma\ \log(1-r_i) \qquad (3.3)$$

with summation running over the smallest roots r_i. For instance, one wants to test the null hypothesis that the rank r is n, i.e. the rank deficiency is 0. One has to look at the sum (3.3) over the smallest root only. If LR rejects, one usually proceeds with the null hypothesis "r is n-1" and calculates LR as the sum over the two smallest r_i. Some fractiles of the distribution of the LR statistic have been tabulated by Johansen (1988).

Note that the LR statistic relies on correlations between differences and level series, conditional on lagged differences and is therefore a direct multivariate generalization of the popular Dickey-Fuller statistic for univariate series. The column vectors of β transform X_t into different variable coordinates which have non-zero correlation to their differences. This property is taken as an indicator for stationarity.

In practice, it is sometimes difficult to fix the lag order p. We suggest to increase p gradually until the residuals from step 5 are white noise according to a portmanteau statistic like the Q by Ljung and Box which is displayed automatically by the RATS software package. Note, however, that this decision depends on the cointegrating dimension r. Therefore, some users prefer using the Q of the regressions in step 1. In most cases, this procedure over-estimates the lag order as the error correction terms should help to whiten the residuals.

Contrary to widespread belief, an over-estimation of the lag order is not innocuous. Of course, high lag orders decrease the degrees of freedom but there is a more important point to this. For an example, take a white noise series. The correlation between the series and its differences is 0.5. If a spurious lag is taken into account, the conditional correlation is reduced to 0.33.[4] In the language of our likelihood problem, inserting spurious lags decreases the chance of identifying cointegrating relations and imposes more integratedness on the system. This is particularly important in the presence of small samples and of inhomogeneous lag structures with the Γ_i matrices showing more zero elements [5] with increasing i. Both properties are met in this paper's examples.

As has been already set out, the correct method to handle seasonal patterns is subject to ongoing discussions. Stochastic seasonality is treated in the next section. However, if seasonality is viewed as deterministic and seasonal dummies are introduced, these can be inserted into the system as well as into all regression (steps 1, 2, 5) in a straightforward manner

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \ldots + \Gamma_{p-1} \Delta X_{t-p+1} + \alpha\beta' X_{t-p} + \sum_{i=1}^{4} a_i D_{it} + \epsilon_t$$

JJ have shown that the distribution of the LR statistic in this

---

4 More generally, it can be shown that k spurious additional lags reduce the correlation between levels and differences for the process $y_t = \alpha\, y_{t-1} + e_t$ to $(1-\alpha)/[k(1-\alpha)+2]$. Compare Kunst (1989a).

5 That is, insignificant elements by their t-value. If these are restricted at zero, the outlined procedure is not maximum likelihood and the zero restrictions would have to be imposed on estimation.

case deviates from that in the original setting and corresponds to the case of taking intercepts into account (JJ's Table VI). Most economic series not only are well represented by integrated processes but also show non-negligible "drifts" and, therefore, this corrected distribution is the more relevant one for most empirical problems. A third and also slightly different distribution comes up if intercepts are imposed in the level regression (steps 2 and 5) only or sample averages are subtracted from the individual series before the analysis (JJ's Table VII).

3.2 The seasonally integrated system

The solution to the maximum likelihood problem in the case of seasonally integrated variables is due to Lee (1989) who elaborated in detail the transformations suggested by HEGY (1990) [6]. In principle, the solution is based on a straightforward extension of the analysis of the last section where the spectral pole at frequency zero caused by the unit root at 1 was in the focus of interest. Here, the basic system

$$X_t = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \dots + \pi_k X_{t-p} + \epsilon_t \qquad (3.4)$$

is transformed by an application of the operator that is appropriate for deleting all unit roots, i.e. $\Delta_4 = 1-B^4$, the same way that 1-B has been used to delete the unit root at 1 in (3.2) and thus a direct analogue of the univariate HEGY test in (2.7). Suppose the VAR system is written in its seasonal differences representation ($\Delta_4 = 1-B^4$)

$$\Delta_4 X_t = \Gamma_1 \Delta_4 X_{t-1} + \dots + \Gamma_{p-4} \Delta_4 X_{t-p+4} + \sum_{i=0}^{3} \Gamma_{p-i} X_{t-p+i} + \epsilon_t \qquad (3.5)$$

Application of the seasonal filter to all individual series and VAR modeling of the filtered data is only allowed if $(1+B)(1+B^2)$ cancels from the lag polynomial $\Gamma_{p-3}+\Gamma_{p-2}B+\Gamma_{p-1}B^2+\Gamma_p B^3$. This amounts to imposing $3n^2$ restrictions on the general representation. This becomes more obvious from the decomposition

6 The HEGY paper is dated correctly 1990 in its published version but had been circulating as a discussion paper since 1988.

used by HEGY

$$\Delta_4 X_t = \Gamma_1 \Delta_4 X_{t-1} + \dots + \Gamma_{p-4} \Delta_4 X_{t-p+4} + \sum_{i=1}^{4} A_i Y_{it-p+3} + \epsilon_t \qquad (3.6)$$

The Y_{it} $(i=1,\dots,4)$ are obtained from $(X_t, X_{t-1}, X_{t-2}, X_{t-3})$ via a one-to-one transformation by applying the filter factors $(1+B)(1+B^2)$, $(1-B)(1+B^2)$, $(1-B)(1+B)$ and $B(1-B)(1+B)$ to X_t. If $A_2 = A_3 = A_4 = 0$, the model immediately reduces to the first-order cointegration model of Section 3.1. A_1 then corresponds to the "impact matrix" $\alpha\beta'$.

Writing a system in pure seasonal differences would impose severe restrictions on (3.4). Not only would this imply the absence of cointegration at frequency zero among the seasonally filtered variables, it also entails the existence of n independent sources of seasonal behavior in the system, acting mutually independently at both seasonal frequencies, i.e. semi-annual and annual cycles.

Note the exact parallels to the Johansen representation of Section 2. For example, A_2 can be split up into $\alpha_2\beta_2'$ which makes sense if there is a rank deficiency. The columns of β_2 will contain vectors which remove the root at -1 from the resulting series. $\beta_2'X_t$ will be trending but will not exhibit semi-annual seasonality while $\beta_2'Y_{2t}$ necessarily will be stationary. In other words, the columns of β_2 "cointegrate at frequency π". Similarly, the rows of β_3 from $A_3 = \alpha_3\beta_3'$ cointegrate at frequency $\pi/2$ immediately if $A_4=0$. If $A_4 \neq 0$, properties can become slightly more complicated because of "dynamic cointegration vectors".

A first solution to estimating (3.7) has been elaborated by Lee (1989). Lee's solution suffers from two restrictions: A_4 is generally restricted to be zero and the cointegration vectors at frequency w_1 are estimated independently from eventual restrictions on the structure at frequency w_2. The first assumption excludes eventual polynomial or dynamic cointegrating vectors (PCIV) at frequency $\pi/2$. Little is known about the importance of this restriction in practice.

Lee's method furthermore extensively exploits asymptotic independence between estimates based on $\{Y_{1t}\}$, $\{Y_{2t}\}$, and $\{Y_{3t}\}$ which could impose problems in smaller samples. If this is taken as granted, then β_1 can be estimated from the canonical vectors on $\{Y_{1t}\}$ with respect to $\{\Delta_4 X_t\}$, β_2 from $\{Y_{2t}\}$ and $\{\Delta_4 X_t\}$, finally β_3 from $\{Y_{3t}\}$ and $\{\Delta_4 X_t\}$. Canonical analysis is to be performed conditional on the other $\{Y_{it}\}$ elements (e.g. the $\{Y_{1t}\}$ to $\{\Delta_4 X_t\}$ correlations are calculated conditional on $\{Y_{2t}\}$, $\{Y_{3t}\}$, $\{Y_{4t}\}$) and on p-4 lags of $\{\Delta_4 X_t\}$, the same way that Johansen's procedure uses p-1 lags of $\{\Delta X_t\}$, the original (level) system assumed to be autoregressive of order p.

The corresponding canonical roots can be used to obtain LR_i ($i=1,\ldots,3$) statistics on the rank of the β_i in full analogy to (3.3). Lee (1989) gives significance points for this test. For the frequencies 0 and π, these are very close to the classical Johansen (1988) fractiles. For the frequency $\pi/2$, they are only slightly smaller. The Lee fractiles, however, correspond to homogeneous systems and, as in the case of pure cointegration at frequency zero, change if intercepts are inserted. It is logical to presume that these again will roughly coincide with the numbers given in JJ's Tables VI and VII. Some Monte Carlo experiments of the author corroborate this presumption although it is too early to state this safely. Anyway, in the following JJ's tables will be used for the seasonal cointegration test suggested by Lee. Calculations were simplified greatly due to a GAUSS program code which the author obtained from Siklos (1989). Tables 3a to 3d give the LR_i statistics for the four six-variable country systems.

In detail, the cumulated sums $-T\,\Sigma \log(1-r_i)$ over the smallest roots of the corresponding problem have been used. Under the null hypothesis, these roots are all zero. If the significance points are exceeded, this is taken as evidence that the largest root among them is non-zero which yields the rank of A_i. To make up for the slight differences between tests on ± 1 and on $\pm i$, 10 % fractiles have been used in the latter case instead of the 5 % fractiles in the former case. Values which exceed these bounds have been boldfaced to provide a summary picture.

TABLE 3a: Seasonal cointegration test statistics for the Austrian system [7]

lags	= 0	1	2	3*	4	# roots
w = 0	**217.611**	**128.580**	**125.306**	**154.856**	**201.874**	6
w = π	**137.848**	**113.455**	**121.266**	**122.360**	93.760	
w = π/2	**118.712**	**126.593**	**112.324**	**100.594**	**105.629**	
w = 0	**77.676**	**81.007**	**84.907**	**100.285**	**127.980**	5
w = π	**84.228**	**69.983**	**74.289**	65.364	61.429	
w = π/2	59.387	57.943	61.034	61.521	**65.743**	
w = 0	39.374	45.589	48.358	**58.574**	**74.123**	4
w = π	**56.187**	39.662	41.639	37.870	36.100	
w = π/2	31.230	33.762	35.863	36.459	30.638	
w = 0	19.437	19.401	16.513	24.383	41.383	3
w = π	30.958	19.332	21.141	13.636	14.876	
w = π/2	12.708	13.497	17.991	18.515	11.733	
w = 0	5.469	9.288	7.993	13.483	16.319	2
w = π	10.550	8.503	8.930	4.795	4.012	
w = π/2	3.737	4.068	5.980	6.925	4.770	
w = 0	0.111	3.100	2.129	3.756	2.935	1
w = π	0.142	0.033	0.171	0.021	0.001	
w = π/2	0.233	1.059	1.646	1.098	0.789	

---

7 "# roots" indicates the number of unit roots (first line: 1; second line: -1; third line: ±i) which is tested for. No rejection in the first block e.g. means no cointegration at the respective frequency.

TABLE 3b: Seasonal cointegration test statistics for the German system[8]

lags	= 0	1	2*	3	4	# roots
w = 0;	**236.364**	**121.237**	**137.006**	**149.885**	**146.774**	6
w = π;	**130.537**	**97.432**	**98.720**	76.427	84.110	
w = π/2;	**124.263**	**111.200**	**116.492**	**110.852**	**119.043**	
w = 0;	**107.743**	**72.065**	**88.735**	**98.759**	**82.987**	5
w = π;	**82.258**	46.637	52.154	44.055	52.111	
w = π/2;	51.537	44.132	48.743	52.840	61.183	
w = 0;	**52.457**	42.933	**53.788**	**56.947**	**49.442**	4
w = π;	36.762	22.410	29.554	25.823	32.708	
w = π/2;	24.466	22.677	23.622	22.698	22.430	
w = 0;	24.226	24.883	25.479	**29.888**	24.701	3
w = π;	10.696	10.392	12.275	14.576	14.713	
w = π/2;	9.559	9.947	8.926	9.382	9.527	
w = 0;	4.849	8.282	7.893	8.923	9.870	2
w = π;	5.064	4.970	5.613	5.600	4.840	
w = π/2;	2.329	1.275	0.339	1.540	0.505	
w = 0;	0.002	2.631	1.958	1.239	2.079	1
w = π;	0.734	0.254	0.214	0.312	0.188	
w = π/2;	0.707	0.086	0.000	0.000	0.092	

8 See Table 3a.

TABLE 3c: Seasonal[9] cointegration test statistics for the Finnish system

lags	= 0*	1	2	3	4	# roots
w = 0	**117.007**	**123.574**	**169.396**	**204.146**	**335.202**	6
w = π	96.159	**107.352**	**149.280**	**159.011**	**277.810**	
w = π/2	**118.433**	**114.035**	96.394	**128.694**	**131.502**	
w = 0	**71.552**	**71.985**	**87.832**	**113.774**	**161.548**	5
w = π	60.615	**71.889**	**80.668**	**90.886**	**129.248**	
w = π/2	**70.953**	**67.515**	57.782	**65.232**	**66.055**	
w = 0	40.307	40.010	39.184	**56.128**	**78.801**	4
w = π	34.231	42.589	40.912	**50.706**	**51.849**	
w = π/2	42.111	32.787	32.032	35.490	33.104	
w = 0	18.112	19.901	17.653	23.400	**40.678**	3
w = π	18.425	21.706	20.302	19.476	29.285	
w = π/2	23.074	16.910	15.121	13.914	13.201	
w = 0	6.456	3.467	3.777	8.584	12.921	2
w = π	6.384	8.148	8.929	8.366	12.630	
w = π/2	8.062	5.848	5.971	2.361	3.024	
w = 0	2.102	0.201	0.112	1.859	0.138	1
w = π	0.062	0.634	3.387	3.130	5.643	
w = π/2	0.070	0.304	0.507	0.317	0.876	

9 See Table 3a.

TABLE 3d: Seasonal cointegration test statistics for the British system[10]

lags	= 0	1	2*	3	4	# roots
w = 0	**120.735**	89.230	**106.692**	**124.268**	**132.766**	6
w = π	**143.655**	94.094	87.706	80.491	**102.543**	
w = π/2	**171.508**	**135.497**	**122.086**	**100.996**	74.031	
w = 0	67.652	48.660	**70.530**	**82.797**	**88.729**	5
w = π	**97.300**	64.275	57.989	53.110	54.627	
w = π/2	**93.097**	60.567	63.777	58.803	39.512	
w = 0	30.781	29.587	38.854	**53.844**	**58.872**	4
w = π	**53.652**	39.593	33.746	29.548	31.371	
w = π/2	**59.658**	35.213	36.581	33.390	19.454	
w = 0	15.158	14.928	21.804	27.262	30.745	3
w = π	23.628	17.674	15.027	13.974	15.873	
w = π/2	30.616	15.423	12.881	9.900	6.325	
w = 0	4.398	5.534	7.009	7.161	8.838	2
w = π	11.360	6.856	5.226	5.874	5.258	
w = π/2	10.639	4.598	2.986	1.788	0.853	
w = 0	0.002	0.302	0.377	0.336	0.421	1
w = π	2.741	1.354	1.475	1.111	0.934	
w = π/2	**2.822**	1.131	0.305	0.183	0.068	

10 See Table 3a.

3.3 Empirical evidence on seasonal cointegration

Tables 3a-d report the results from an application of the Siklos (1989) procedure to the multivariate (6-dimensional) systems of the individual countries. It is seen from the tables that cointegration at the seasonal frequencies can be quite substantial and in many cases the correlation coefficients even exceed those at frequency zero, i.e. evidence on seasonal cointegration is stronger than usual frequency-zero cointegration.

With regard to lag order specification, up to four lags of $\Delta_4 X_t$ were tentatively inserted. Some lines, however, exhibit "perverse" behavior as the correlation between stationary and non-stationary variates *increases* as the lag order is extended. It is obvious that the true conditional correlation must decrease if the conditioning set becomes larger. If the sample correlation increases this is a strong indication for an exhaustion of the sample information by relative overparametrization. Kunst (1989a) reports that this effect is felt strongly for more than around ten lags in a sample of 100 observations on a univariate random walk. In the multivariate system which is treated here, the effect could be troublesome for fewer lags already.

It is difficult to give a definite answer to the question which lag order can be regarded as the correct one. It seems, however, that the Austrian system demands for an augmentation of 3 lags to render uncorrelated vector residuals and that the Finnish system does not need any augmentation (0 lags). In the cases of Germany and the United Kingdom, 2 lags could be a good choice but leave substantial residual autocorrelation in some directions. Particularly for the United Kingdom, the C series would demand for an excessive number of lags of C and the other variables to be described with clean errors in an unrestricted vector autoregression. Suggested lag orders are marked by asterisks in the tables.

Numbers corresponding to the frequencies 0 and π in Tables 3a-d are marked in bold face if they are significant at the 5 % level according to JJ's Table VI. At the frequency $\pi/2$, values are boldfaced if significant at 10 %. According to this criterion, at least one seasonal cointegrating vector exists for all countries. A second seasonal vector is supported for Finland. If this result really holds, it is not surprising. If exports are non-seasonal, which is the case for most countries, maybe with the exception of Germany, one vector is given by the corresponding unit vector. If no seasonality comes in from the inventory changes, the state sector, and imports, one is tempted to conclude that seasonality in national income Y evolves from seasonality in consumption C and investment I, which establishes a second seasonal cointegrating vector.

The analysis of the eigenvectors is to be used with care only. In some cases, the vectors corresponding to the largest eigenvalues of the A_1 problem, for example, were unable to cointegrate the seasonally filtered series convincingly which they are supposed to do according to the procedure. We shall return to these vectors in the next section. In much the same way, the eigenvectors corresponding to the A_2 and $A_3 \mid A_4$ problem did not always extract the kind of seasonal cycles which they were supposed to annihilate. Possible explanations for this phenomenon are manifold: overparametrization, ignoring of residual correlations due to modeling ARMA structures by pure VARs, deviations from distributional assumptions, cross-effects among the Y_{it} processes in smaller samples, near-integration of second order. The general dilemma of the algorithm, however, seems to be that one instinctively searches for maximum *unconditional* correlations although this is clearly done via calculating *conditional* correlations. Theoretically, a series with non-zero conditional correlations cannot display zero unconditional ones but, sadly, empirical results sometimes point in this direction.

In the following, the first two canonical vectors or eigenvectors corresponding to the semi-annual frequency will be described for each country. These are denoted v_1 and v_2, v_1 corresponding to the largest root, v_2 to the second largest one. Later, we will use the remaining v_i in analogous notation, i.e. v_6 corresponding to the smallest root etc. With the German data, these vectors v_1 and v_2 which are supposed to cointegrate at frequency π (only the first one is significant) are given approximately by the following vectors (coefficient on R very small and suppressed):

v_1 = Y+0.78*C+0.42*I-1.27*X-1.38*W

v_2 = Y-0.48*C+0.32*I+0.28*X-0.68*W

Here, v_1 could be interpreted as relating seasonality in wages to the seasonal structure of the national account aggregates Y, C, I with an interesting adverse contribution from exports. Using some algebraic transformations, a linear combination of v_1 and v_2 can be viewed as explaining seasonality in output (Y) primarily by seasonality in C and X. Furthermore, it is seen that the X unit vector is not contained in the linear space generated by v_1 and v_2 which is further evidence on the seasonal nature of German exports.

A similar experiment performed on the British data yields the following two vectors:

v_1 = 1.80*Y+0.35*C-0.39*I-2.45*X+0.32*R+1.35*W
v_2 = 5.91*Y-5.77*C+2.73*I-0.01*X+1.25*R+4.67*W

The second vector eliminates seasonality in the consumption quota C-Y by help of W while the first one relates seasonal exports to output and wages. The comparatively small influence of I is due to the weak seasonal pattern of that variable (compare Figure 3 and Table 2b). This weakness of the seasonal investment pattern separates the United Kingdom from the other countries which suffer from the influence of cold winters impairing investments in the construction sector. Surprisingly, non-seasonal R again fails to show big coefficients in the vectors. [11]

For Austria, the following vectors were obtained:

v_1 = Y-0.01*C-0.25*I-0.01*X-0.01*R-0.06*W
v_2 = Y-0.11*C-0.45*I-0.39*X-0.07*R+0.60*W

Again, R does not enter in both vectors. The statistically significant v_1 seems to relate the seasonality in Y to I whereas

11 Scales in the series R and W have been adjusted in order to have the same magnitude as the accounts series Y,C,I.

the remaining variables contribute very little to seasonal (semi-annual) output cycles. The less significant v_2 has an easier interpretation if v_1 is subtracted. After doing so, it relates the seasonal pattern in W to the demand components C, I, and X, the appearance of the latter one being interesting as it looks non-seasonal at first sight.

Finally, the Finnish data yield the following two vectors that should cointegrate at frequency π:

$$v_1 = 5.07*Y-11.98*C-0.46*I-0.41*X+1.75*R+8.87*W$$
$$v_2 = 1.65*Y+0.75*C-1.26*I+0.91*X-0.19*R-3.70*W$$

The first vector links seasonality in C to influences from output, from wages, and the interest rate, the latter two series also reflecting seasonal cycles in the price deflator due to their construction. v_2 treats seasonality in I and links it to wages and output. v_1 shows a rather strong influence of R which can be interpreted as the influence of short-run interest rate fluctuations on C while the absence of R in v_2 could indicate that these fluctuations have no impact on investors' behavior.

Figure 7a-d provide for graphical displays of the sample autocorrelation functions (ACF) of all six components generated by the system rotations which are given by the vectors v_1, v_2, and the remaining v_3 to v_6. The generated components have been filtered by $1-B+B^2-B^3 = (1-B)(1+B^2)$ to get rid of the remaining unit roots whose elimination the canonical problem on the factor 1+B has not been designed for. Notwithstanding eventual distorting effects caused by the conditioning variates, one expects from the ACF graphs that they are ordered such that the first one is certainly stationary, that the sixth one clearly reflects the unit root at -1 and that the remaining ones are somehow in between. This expectation is satisfied by Figure 7.

Instead of the ACF, the spectrum could have been used as a visual criterion but the ACF was preferred here for two reasons: firstly, it is conveniently standardized whereas spectra allow for a variety of versions according to window size, window shape, and normalization of the ordinate axis; secondly, the ACF is less prone to produce arbitrary wiggles that fool the eye in smaller samples as all the shown points are natural parameters.

Returning to Figure 7 in more detail, the following basic shapes can be distinguished:

1) The ACF of a white noise series displays only arbitrary and insignificant deviations from the abscissa axis.

1a) The ACF of a stationary series with reasonably short memory shows significant deviations at the first few lags only or a quick decay.

1b) The ACF of a series over-differenced by $1-B+B^2-B^3$ starts with a distinct 1/minus/plus/minus pattern and then parallels 1a or 1b.

2) The ACF of a series containing 1+B shows a repetitive and only slowly decaying zigzag pattern.

Three of the German and British ACFs point to the class 2 but only two of the Finnish ones and only the last one in the Austrian case. Recognizing that the Lee-Siklos testing procedure tends to categorize borderline cases into the second class due to the principle of conservative testing, visual evidence gives substantially more "cointegrating" vectors than formal testing, classifying a substantial share of the transformed series into category 1b, e.g. the first British component which, however, shows disturbing wiggles around lag 20 which might be responsible for its failure to cointegrate formally. The only surprising feature found in these graphs is that Germany seems to have substantially more independent sources of stochastic season than Austria. The concentration of all seasonal structure into the last component could even enhance the possibility of a completely deterministic model of seasonality for Austria whereas this is obviously impossible in all remaining cases.

A comparison of Figure 7 with the univariate statistics of Table 2, particularly t_2 which tests for 1+B, allows further conclusions. In Austria, the only safely stochastic fluctuation at this frequency in the consumption series and the deterministic fluctuations are contracted in the last component, the rest of the system being free from seasonal patterns. There is only one source of stochastic seasonality, possibly consumption, which in turn may influence other aggregates - and therefore, literally, much "seasonal cointegration", all words like "season" meaning just the semi-annual frequency. Over-differencing effects and the high number of conditioning terms (a lag order of 7) prevent the LR statistic from rejection.

In Germany, an economy with clear evidence on stochastic seasonality in most series, there are three independent sources of that phenomenon, a fourth one (ACF_3) being weak and unconvincing, and therefore three or at least two true and meaningful seasonal cointegrating vectors. In Finland, the shortest data set of the four, ACF_3 and ACF_4 leave the spectator without a decision. Two sources of semi-annual season are certain, one of them "engulfing" the deterministic season, and two vectors of seasonal cointegration at the other end of the scale.

In the United Kingdom, the country where deterministic seasonality is most prevalent, the stochastic sources in the consumption and income series and the deterministic shapes are mirrored by the ACF plots which leaves virtually no room for any seasonal cointegration effects.

Restricting attention to v_5 and v_6, i.e. the vectors generating the most virulent sources of stochastic season, enables a comparison among the four countries. By calculating mutual correlations, it is immediately seen that Austria, the United Kingdom, and Finland share one very similar source of seasonality which is not present in Germany and is almost identical to Y-0.8*C, among the remaining entries only X being more conspicuous. Germany and Finland share another vector which is related to seasonal fluctuations in the wage and output series (weights of equal sign, see below). In contrast, the possibly cointegrating vectors v_1 and v_2 vary notably among countries.

We now turn to the seasonal cointegration problem at frequency $\pi/2$ or, equivalently, at the factor $1+B^2$ or at the complex unit roots $\pm i$. The search is restricted to static vectors ($A_4=0$) and thus it may ignore some cointegrating structures which need one lag in the variables. An eventual underestimation of seasonal cointegrating structures entails an eventual overestimation of the number of independent sources of annual cycles in the system. Again, the ACF of all related components are displayed in Figure 8a-d. The components have been filtered by $1-B^2$ to get rid of the roots at ±1. A checklist similar to the one from the semi-annual problem can be constructed easily.

It is seen from Figure 8 that, not unexpectedly, independent seasonal cycles are fewest in the United Kingdom, possibly only two. Austria and Finland show evidence on a third non-stationary component whereas Germany displays four seasonal components. Again, there seems to be substantially more cointegration in the system than can be seen from the statistics in Table 3.

The tentative interpretation of the vectors at frequency $\pi/2$ should be more succinct than that at frequency π as these are less reliable due to the restriction of being static. Both frequencies usually come together in economic series and it is difficult to decide which of the two frequencies is the more important one. The vectors corresponding to the two largest roots are given in Table 4. Although the individual vectors do not coincide across countries, common vectors are shared approximately by the spaces spanned by v_1 and v_2. To begin with, Austria and Germany share a three-entry vector Y-1.3*X-.4*W which indicates that the seasonal structure of the exports series of these countries can be eliminated by accounting for wages and output. These countries are neighboring countries with strong economic interdependencies and similar climatic conditions and, therefore, this vector could even make sense economically. Secondly, Austria and the United Kingdom share a vector Y-0.6*C with smaller entries by the other variates, including interest. This vector relates the seasonal fluctuations in consumption to the weaker ones in output by appropriate weighting. More interesting than the appearance of this vector in two non-related economies is its failure to appear (or at least of a similar vector) in Germany and Finland. Finland, however, has a common vector with the United Kingdom which links consumption to the wage-output ratio taking the remaining variables into account via lesser coefficients, i.e. Y-0.8*C-W±... This means that seasonality in C is best explained by Y in Austria but by W-Y in Finland, the United Kingdom allowing for both interpretations.

At the other end of the scale, the main sources of the annual fluctuations are given by the last two components and thus by v_5 and v_6. Here, Germany and Austria share a structure close to Y-0.5*C-.0.3*I+0.1*R+0.2*W, the coefficients in the other variables serving to enhance seasonality in Y. Germany and Finland are connected by a simpler structure, W-0.8*Y with very small entries by other variables. Table 2 corroborates the view that seasonality is strong in German as well as Finnish wages. At first, it is surprising that the non-stationary variates do not have entries of the same sign from W or Y at the same time but this is only an indication how one tends to mix up integration at frequency 0 and integration at cycles: W+Y certainly is strongly trending but it is not necessarily more strongly seasonal than combinations like the German-Finnish one. Yet more confusingly, the sign of Y and W is the *same* in the common seasonal component vector of Germany and Finland at π but the signs alternate at $\pi/2$.

Lee (1989) also treats the special case that the same vectors cointegrate at both seasonal frequencies, which he calls "uniform seasonal cointegration", and develops a test for this hypothesis. At first sight, this property does not appear to be present in any of the four systems, another indication that the true cointegrating vectors might be dynamic. The only economy coming close to uniform seasonal cointegration is the United Kingdom with a cointegration vector relating C-Y to wages (positively) and maybe interest. Even there, it is a linear combination of v_1 and v_2 that does the trick, the vectors v_1 being completely different at the two frequencies. Anyway, Y-0.8*C is recognized as a source of seasonal fluctuations *both* at the π and at the $\pi/2$ frequency in Finland and there is a somewhat weaker correlation among vectors relating investment to wages at both seasonal frequencies in Germany whereas no such correspondences can be found in Austria or the United Kingdom.

TABLE 4: First two solution vectors to the seasonal cointegration problem at frequency $\pi/2$

	Y	C	I	X	R	W [12]	
Austria							
	1.000	-0.226	-0.188	-0.601	0.737	-0.127	*
	1.000	-0.548	-0.146	0.166	0.988	0.129	
Finland							
	1.000	2.453	-0.763	-0.022	1.220	3.040	*
	1.000	-0.112	-0.334	0.430	-0.053	-0.146	*
Germany (Federal Republic)							
	1.000	0.701	-0.469	-1.980	-0.063	0.217	*
	1.000	-0.443	0.358	-0.431	0.619	-0.962	
United Kingdom							
	1.000	-0.657	-0.187	-0.673	1.273	2.651	*
	1.000	-0.713	-0.114	0.186	0.111	0.118	

12 The numbers beneath the labels denote the coefficients of: gross national (or domestic) product; private consumption; gross fixed investment; real interest rate; exports; real wages. Significant vectors are marked by asterisks.

3.4 Cointegration at frequency 0

A detailed analysis and interpretation of the long-run structures encountered in the four country systems is given in a companion paper [Kunst (1990)] but the cointegrating vectors identified by the algorithm are repeated in Table 5 for the sake of completeness. The displayed vectors are based on the model specifications starred in Tables 3a-d. Note that, according to the statistics, three vectors cointegrate in Austria and Germany but only two in Finland and in the United Kingdom.

TABLE 5: Cointegrating vectors at frequency zero for selected country models

	Y	C	I	X	R	W [13]
Austria						
	1.00	-.29	-.28	-.22	.01	.07
	1.00	-.54	.64	-.12	.32	-.68
	1.00	.10	-.10	-.35	.02	-.17
Finland						
	1.00	.51	-.88	.08	-1.05	-1.82
	1.00	-.63	-.28	-.27	.81	.12
Germany, Federal Republic						
	1.00	-.64	-.30	-.14	-.14	.18
	1.00	.44	-.03	-.43	-.04	-.54
	1.00	.30	-.85	-.43	.48	.24
United Kingdom						
	1.00	1.01	-1.98	-.36	-.02	-.06
	1.00	-.56	-.18	-.35	.00	.45

---

13 The numbers beneath the labels denote the coefficients of: gross national (or domestic) product; private consumption; gross fixed investment; real interest rate; exports; real wages. Significant vectors are marked by asterisks.

3.5 The alternative model : deterministic seasonality

After this detailed study of all structures encountered on the basis of stochastic seasonality and seasonal unit factors, it may be interesting to see how things change if deterministic seasonality is assumed and therefore seasonal dummies are inserted into the regressions. Taken literally, such a system is suspect as now seasonality is granted too much freedom. It means that the seasonal pattern is allowed to persistently change its shape and to expand at the same time. Therefore, no quantitative results of this experiment are tabulated in this paper.

If the deterministic model is correct, one would expect the following phenomena: firstly, the optimum lag order of the VAR will decrease as the stochastic seasonal model entails over-differencing and artificial moving-average terms; secondly, all roots on the π and $\pi/2$ will be significant as the analysis is conditioned on the dummies and the residuals are non-seasonal; thirdly, the roots at frequency zero, particularly the largest, will not change too much, otherwise (e.g. with roots of 0.6 to 0.7) it would indicate seasonal unit roots.

In Section 2, we have seen already that the degree of stochastic fluctuations in seasonal patterns differs among countries. Germany seems to have more stochastic seasonality than the United Kingdom. These findings are corroborated if the seasonal cointegration procedure is applied conditional on seasonal dummies.

In the case of Germany, the optimum lag order *increases* from 2 to 3 relative to the purely stochastic model. There is now evidence on four cointegrating vectors at $\pi/2$ but the evidence remains unclear on π. The number of cointegrating vectors at 0 increases to 4. In summary, insertion of dummies is not enough to get rid of the seasonal unit root effects in the system.

In the case of the United Kingdom, the optimum lag order decreases to 0 although some doubts remain with respect to the consumption series. All canonical roots at π and $\pi/2$, including the smallest one, are significantly different from zero but this observation ceases to hold if two or more lags are inserted. All numbers related to the zero frequency are replicated almost exactly from Table 3d. In summary, British seasonality could be entirely deterministic while German seasonality is, at least to a certain degree, stochastic.

The two smaller economies produced intermediate results between the two extreme cases of Germany and the United Kingdom.

4. Summary and conclusions

It is difficult if not impossible to give definite answers on behalf of questions of classification such as integrated versus trend-stationary on the basis of finite samples. The same goes for the classification of the seasonal patterns. Properties at the seasonal frequencies such as integration or cointegration are inseparable from near-cointegration at these frequencies, the same way that integration at zero is inseparable from near-integration at low frequencies. Perhaps, it makes more sense to speak of vectors that are able to reduce the correlation structure of the resulting series substantially, compared to the components series, a point which has been taken up under the name of "codependence" in the French econometrics literature (compare Gourieroux and Peaucelle (1988)). This would also be compatible with the viewpoint that the question of classification has been given too much weight recently which has also been addressed by Christiano and Eichenbaum (1990) and has given rise to a closer look at possible intermediate structures such as fractionally integrated models (see e.g. Geweke and Porter-Hudak (1983)).

One more caveat must be attributed to the treatment of seasonality. From the results of this paper as compared with related studies [Kunst (1989b,1990)], it is seen that the selected procedure to extract seasonality influences the results quite substantially. The influence of the usual univariate "seasonal adjustment" on the estimation and testing in the cointegration framework is still an open field for research. Jaeger and Kunst (1990) have shown the strong influence of Census X-11 on univariate measures of integratedness and the influence on multivariate measures, such as the LR statistics used in this paper, may even be more troublesome.

As far as evidence on the phenomenon of seasonal cointegration is concerned, this paper's results corroborate the presumption that it may be quite common in multivariate systems for several reasons. Firstly, all series which do not have seasonal unit roots univariately - even if they should show signs of weak or deterministically stable seasonal patterns - are by definition cointegrated with themselves, i.e. the corresponding unit vectors cointegrate. Secondly, the absence of seasonal cointegration would mean that there are as many sources of seasonality as series which is, again by definition, impossible if output, consumption, and investment are included in the system and the remaining aggregates (i.e. output minus private consumption and investment) is non-seasonal. Compared to this, the numbers of seasonal cointegrating vectors identified by the test procedure alone are surprisingly small but the analysis has shown that these are underestimated in some cases.

It remains to investigate whether these more complete models are actually helpful, for instance, in increasing forecasting precision. In many cases, although this of course depends on the tradition set by the national statistical offices, quarterly forecasts on unadjusted series and annual forecasts are more interesting than quarterly adjusted values. Additionally, if cross-effects as strong as seasonal cointegrating vectors are present, even the annual forecasts should be improved by using the full model.

Note that, in the presence of seasonal cointegration, some popular modeling strategies are necessarily incorrect, in particular: individual seasonal adjustment by Census X-11; individual seasonal filtering by four-quarter moving averages; VAR modeling with seasonal dummies included. Interestingly, the deficiencies of the latter strategy are quickly seen from the *largest* roots of the cointegration test which, in that case, exceeds 0.5 and corresponds to a strongly negatively autocorrelated series, i.e. the component which contains the unit root at -1. This phenomenon can be obtained by applying the procedure to *any* (including U.K.)

of the four country systems of this study, another indication for the importance of stochastic seasonality.

Acknowledgments

Credits to Sören Johansen and Marius Ooms for helpful comments on a previous version of this paper. Credits to Klaus Neusser for the German data and the basic selection of the economic variables. Credits to Pierre Siklos for the algorithm for the seasonal cointegration analysis. All errors are mine.

Literature:

Box, G.E.P., and G.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control*. 2nd edition, Holden-Day, San Francisco.

Christiano, L.J., and M. Eichenbaum (1990) "Unit Roots in Real GNP - Do We Know and Do We Care ?". In: Unit Roots, Investment Measures and Other Essays, *Carnegie-Rochester Conference Series on Public Policy* 32, 6-46.

Dickey, D.A., and W.A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series With a Unit Root". *Journal of the American Statistical Association* 74,427-431.

Engle, R.F. and C.W.J. Granger (1987) "Co-integration and error correction : Representation, estimation and testing". *Econometrica* 55, 251-276.

Engle, R.F., C.W.J. Granger and J.J. Hallman (1989) "Merging Short- and Long-Run Forecasts: An Application of Seasonal Cointegration to Monthly Electricity Sales Forecasting". *Journal of Econometrics* 40, 45-62.

Engle, R.F., C.W.J. Granger, S. Hylleberg, and H.S. Lee (1990) "Seasonal cointegration: The Japanese consumption function 1961.1-1987.4". Paper presented at the Econometric Society

World Conference, Barcelona, Spain.

Geweke, J. and S. Porter-Hudak (1983) "The estimators and application of long memory time series models". *Journal of Time Series Analysis* 4, 221-238.

Gourieroux, C. and I. Peaucelle (1988) "Detecting a Long Run Relationship (with an Application to the P.P.P. Hypothesis)". CEPREMAP Research Memorandum N-8902, Paris, France.

Hylleberg, S., R.F. Engle, C.W.J. Granger and B.S. Yoo (1990) "Seasonal Integration and Cointegration". *Journal of Econometrics* 44, 215-238.

Jaeger, A. and R.M. Kunst (1990) "Seasonal Adjustment and Measuring Persistence in Output". *Journal of Applied Econometrics* 5, 47-58..

Johansen, S. (1988) "Statistical analysis of cointegration vectors". *Journal of Economic Dynamics and Control* 12,2/3,231-254.

Johansen, S. and K. Juselius (1989) "Hypothesis Testing for Cointegration Vectors for Gaussian Vector Autoregressive Processes with Constant and Seasonal Dummies". Working paper, University of Copenhagen.

Kunst, R.M. and K. Neusser (1990) "Cointegration in a Macro-Economic System". To appear in the *Journal of Applied Econometrics*.

Kunst, R.M. (1989a) "A Note on the Influence of Order Selection on Correlation Analysis in Autoregressive Processes". mimeo, Institute for Advanced Studies, Vienna.

Kunst, R.M. (1989b) "Cointegration in a Macro-Economic System: Seasonality and Explosive Roots". *IAS Research Memorandum* No. 255, Institute for Advanced Studies, Vienna.

Kunst, R.M. (1990) "Cointegration in Macro-Economic Systems: Case

Studies for Small and Large European Economies". mimeo, Institute for Advanced Studies, Vienna.

Lee, H.S. (1989) "Maximum Likelihood Inferences on Cointegration and Seasonal Cointegration". Discussion paper, University of California, San Diego.

Siklos, P. (1989) Personal communication.

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.

Tso, M.K.-S. (1981) "Reduced-rank Regression and Canonical Analysis". *Journal of the Royal Statistical Society B* 43, 183-189.

FIGURE 1: Gross domestic (national) product series

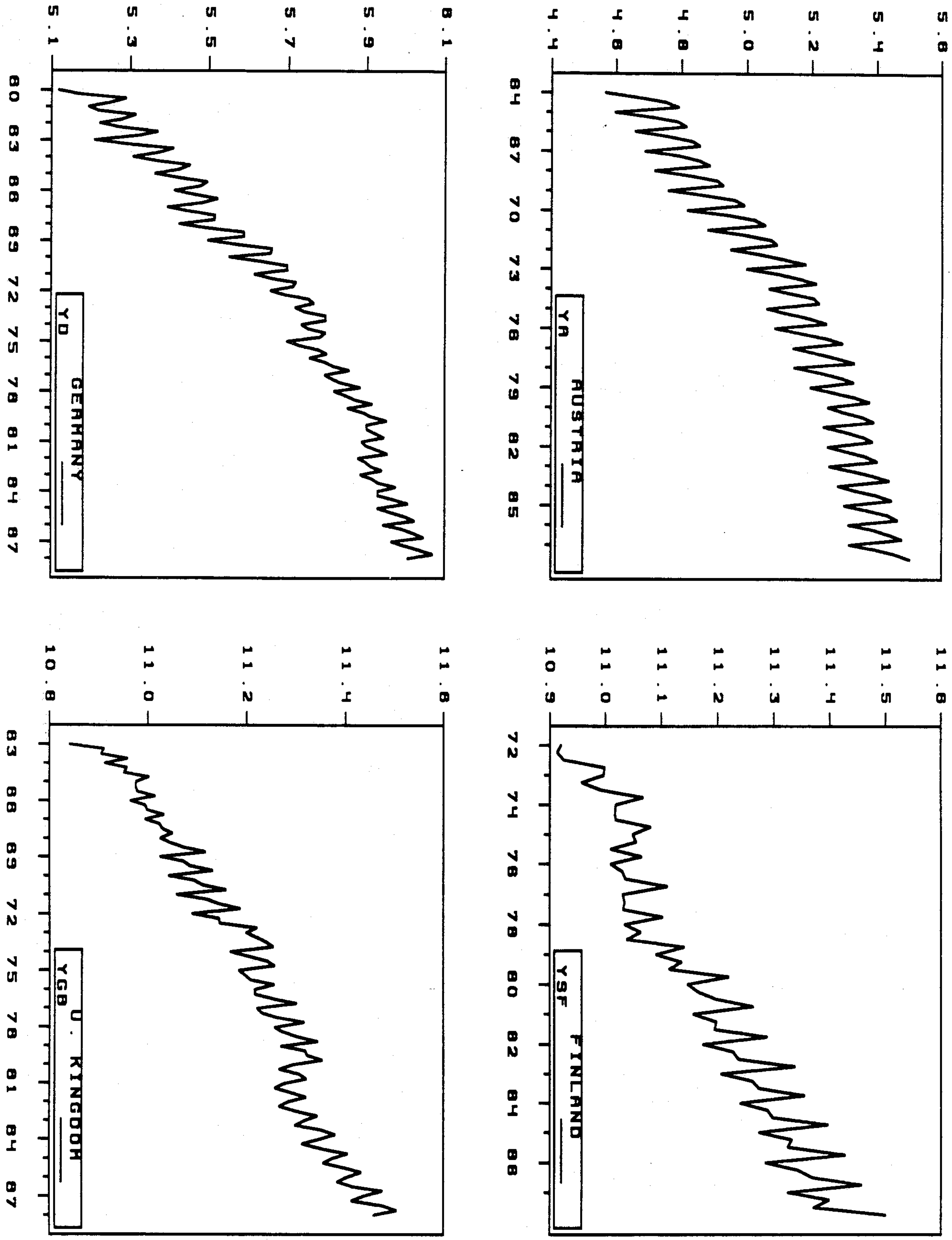

FIGURE 2: Private consumption series

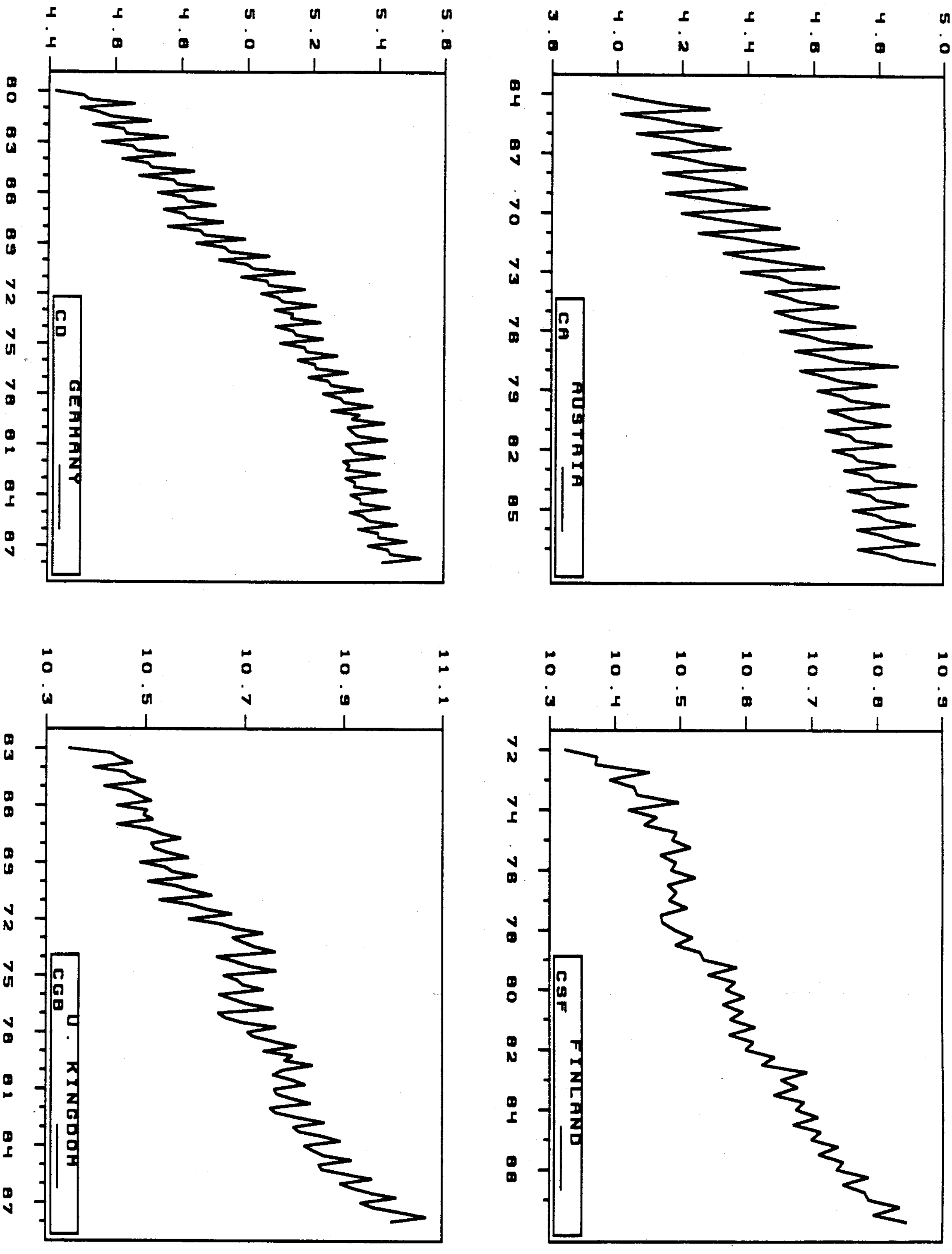

FIGURE 3: Gross fixed investment series

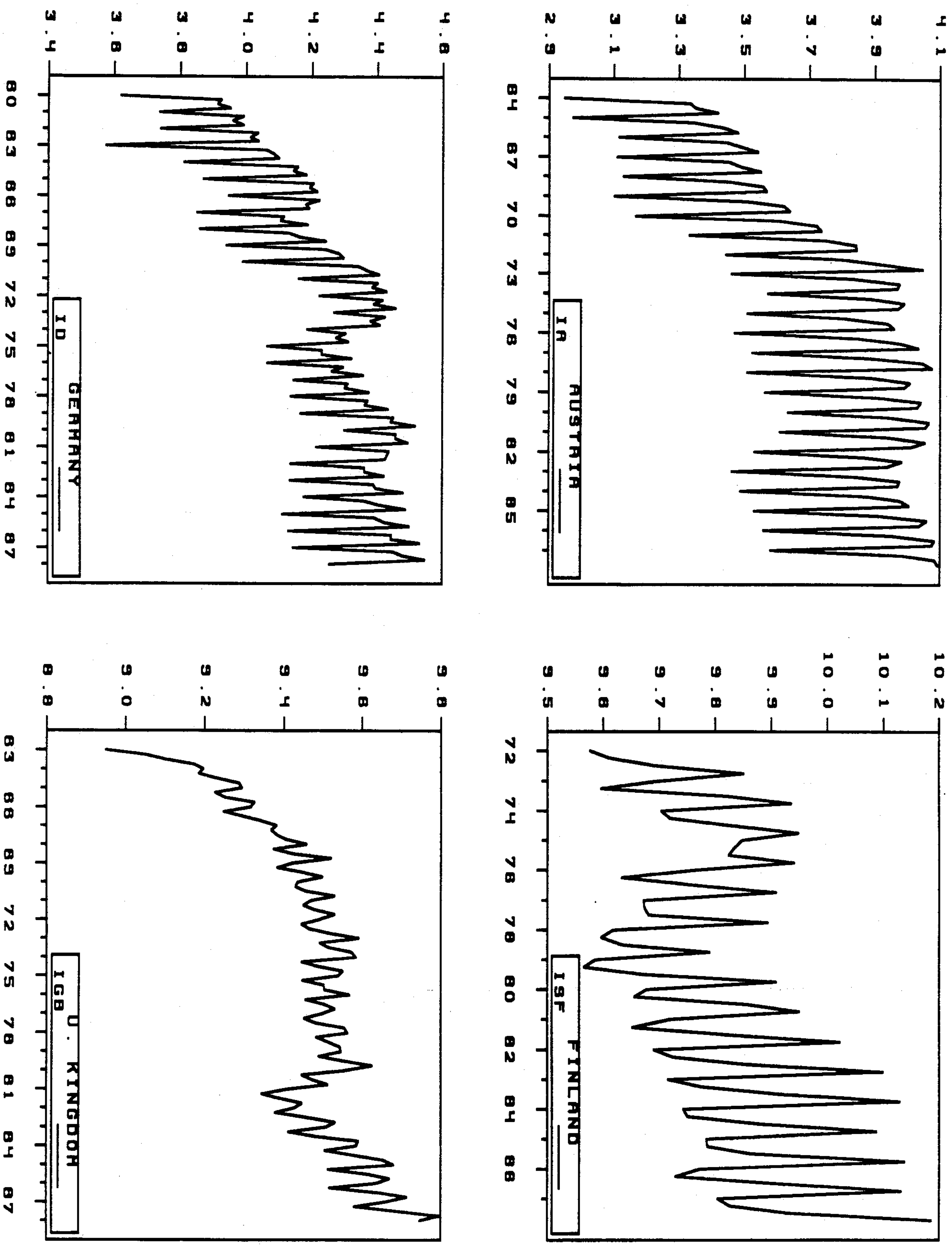

FIGURE 4: Exports series

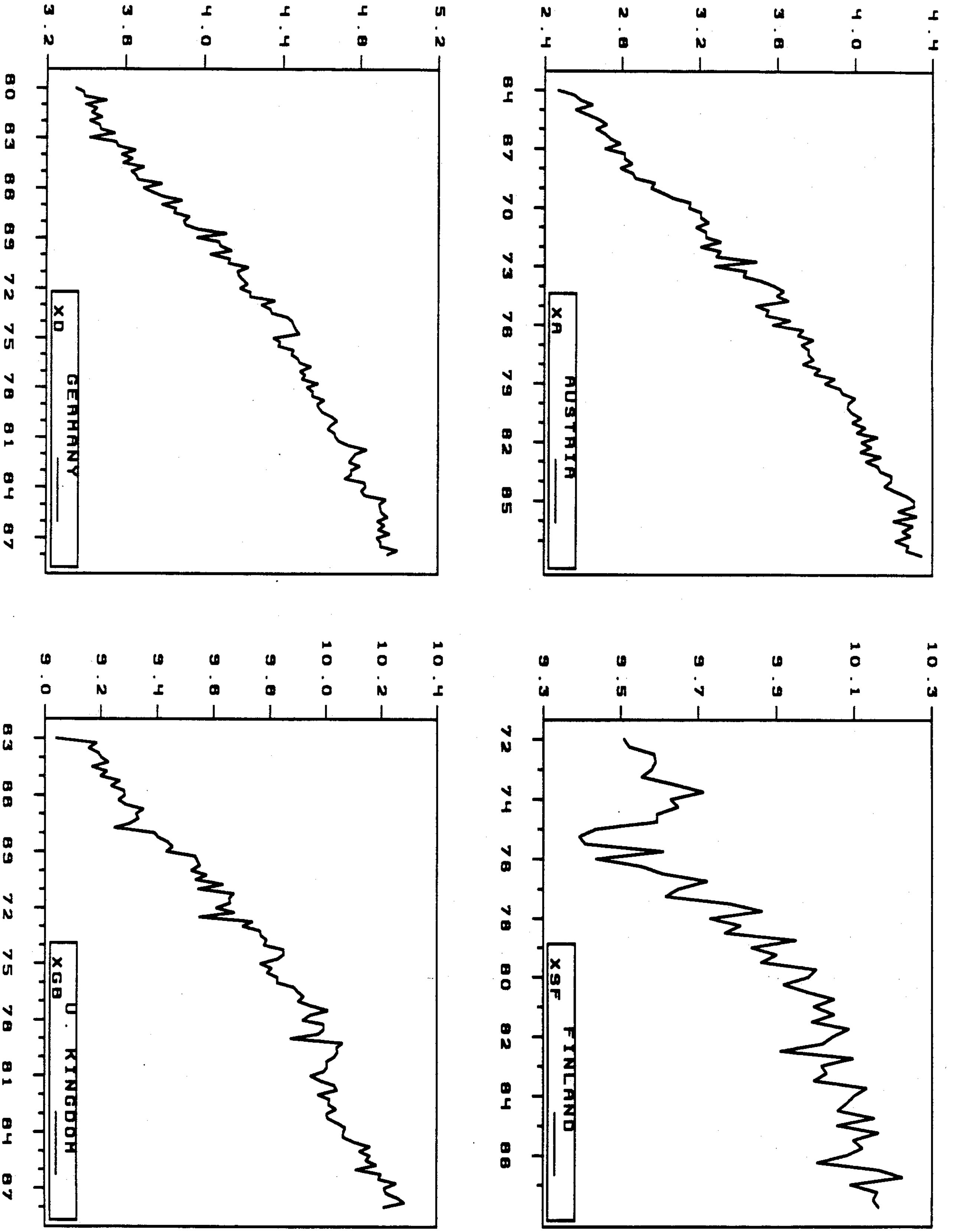

FIGURE 5: Real interest rate series

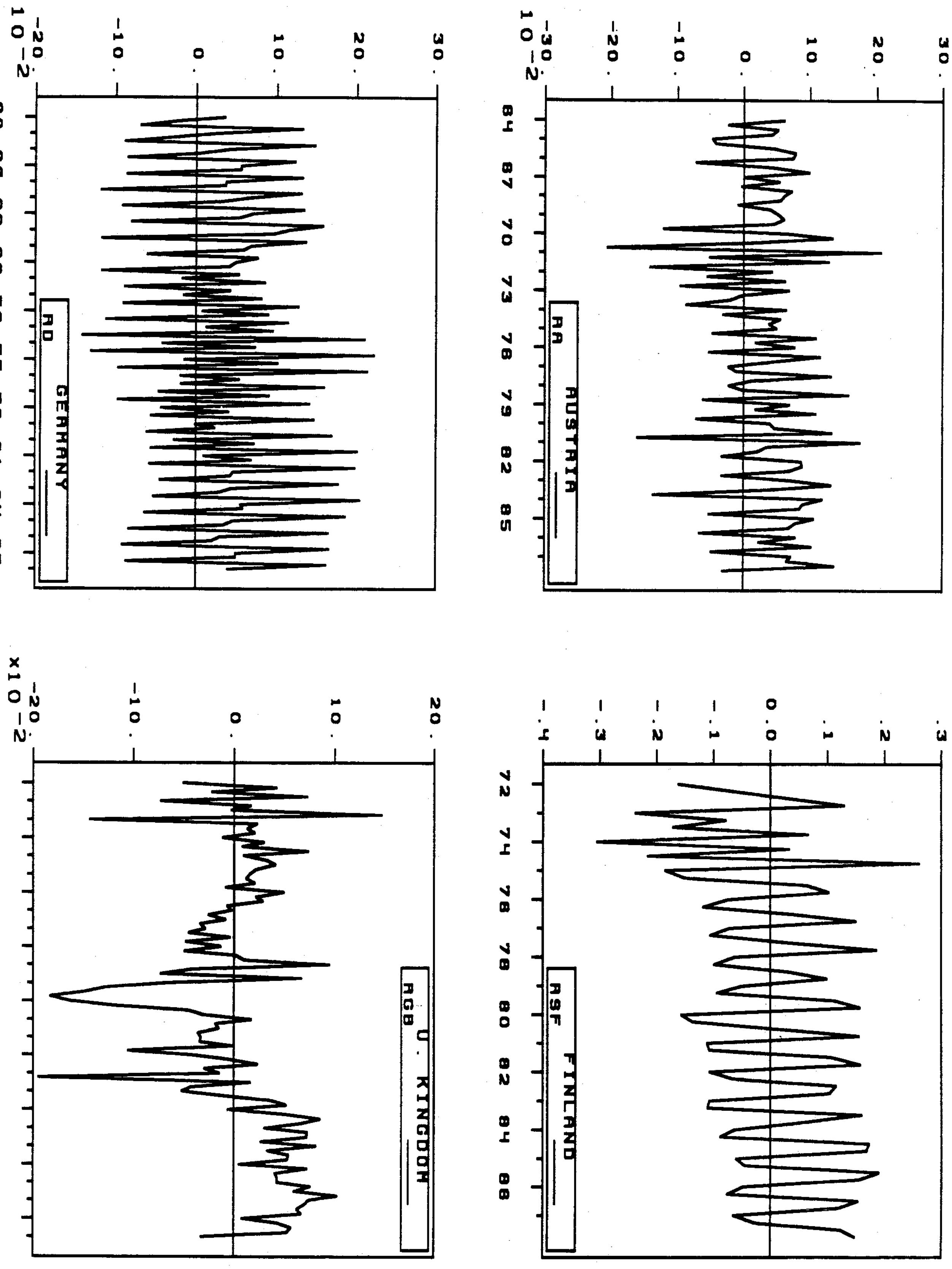

FIGURE 6: Wage series

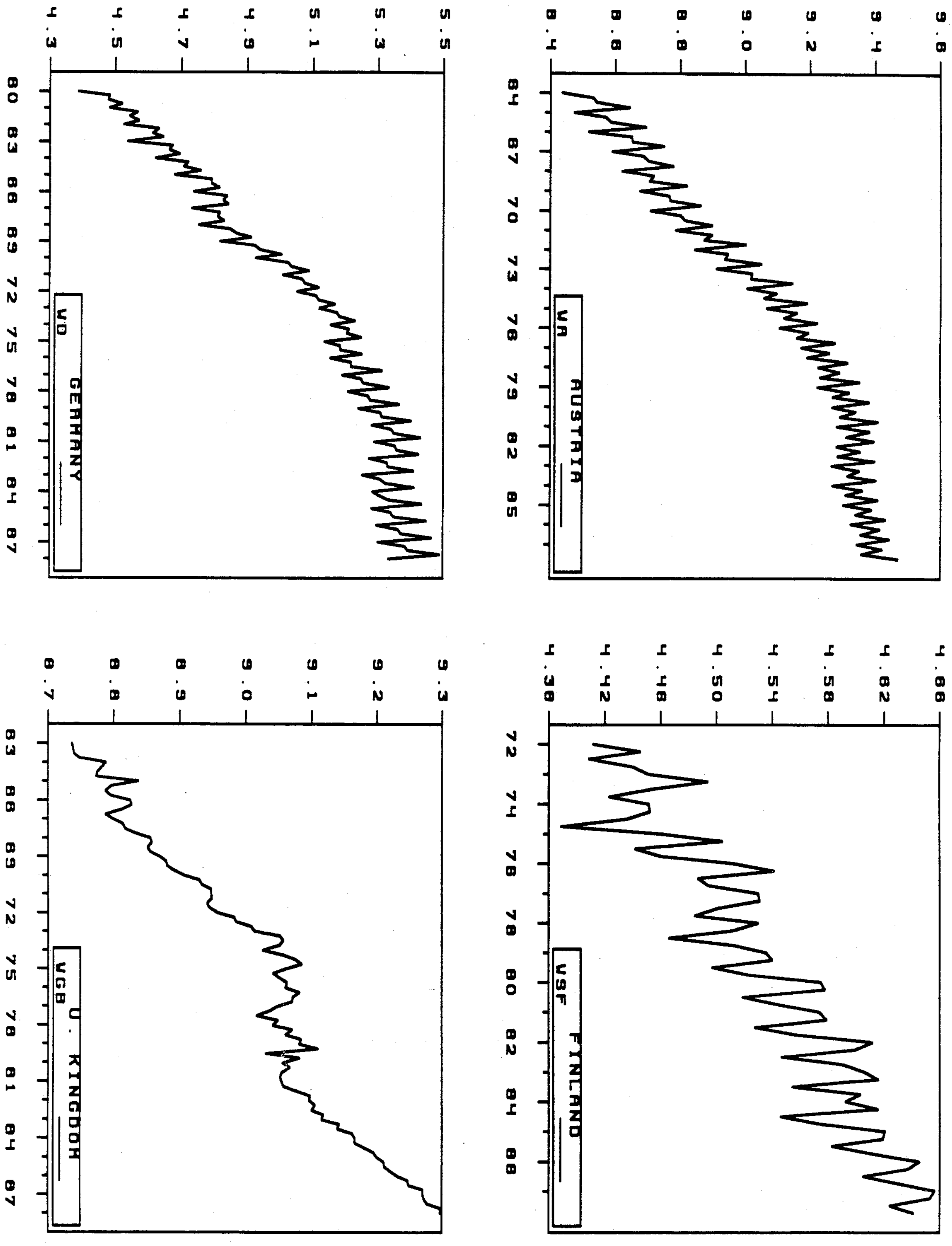

FIGURE 7a: Autocorrelation functions of the six components of the Austrian system which emerge as solutions of the seasonal cointegration problem at frequency π. All components filtered by $1-L+L^2-L^3$

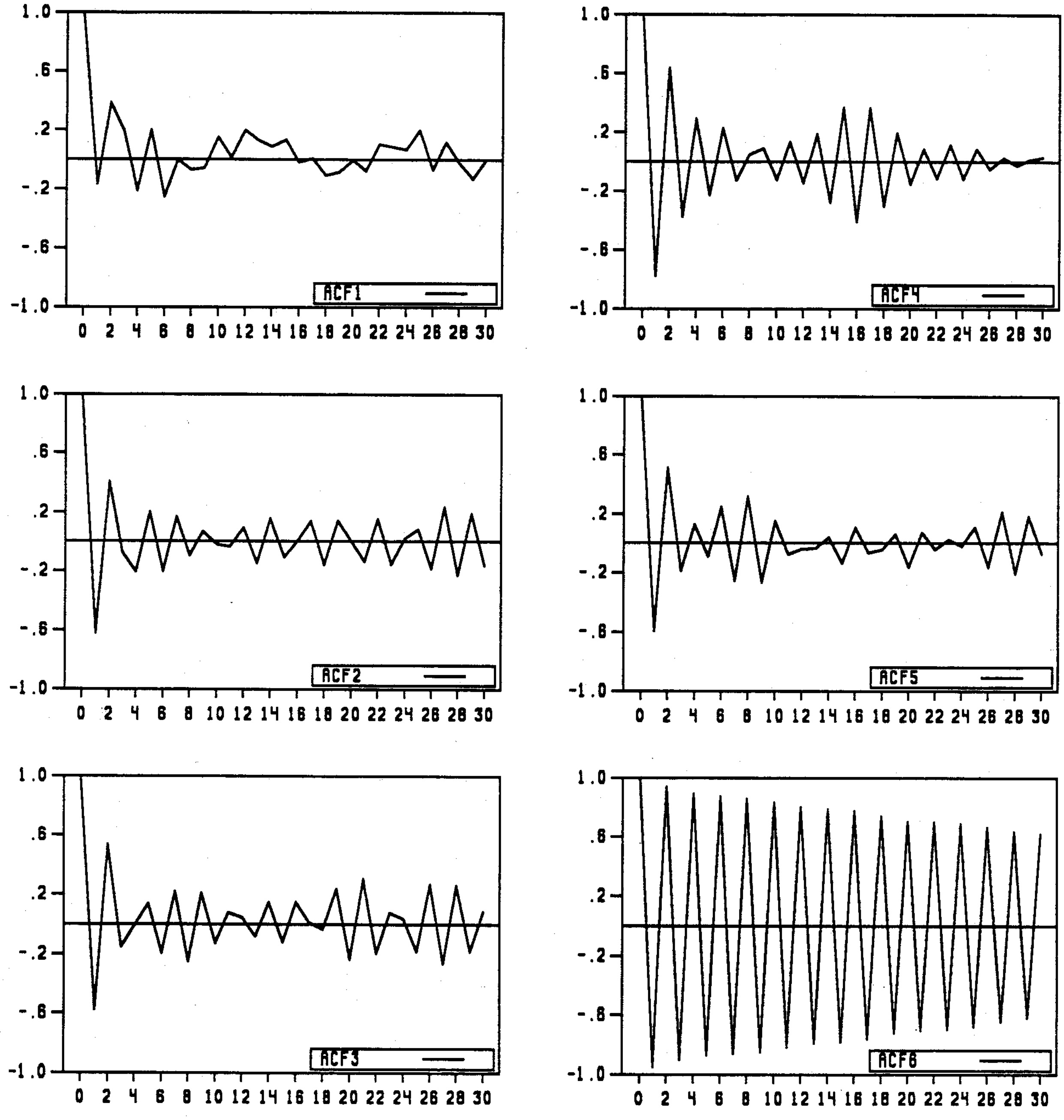

FIGURE 7b: Autocorrelation functions of the six components of the German system which emerge as solutions of the seasonal cointegration problem at frequency π. All components filtered by $1-L+L^2-L^3$

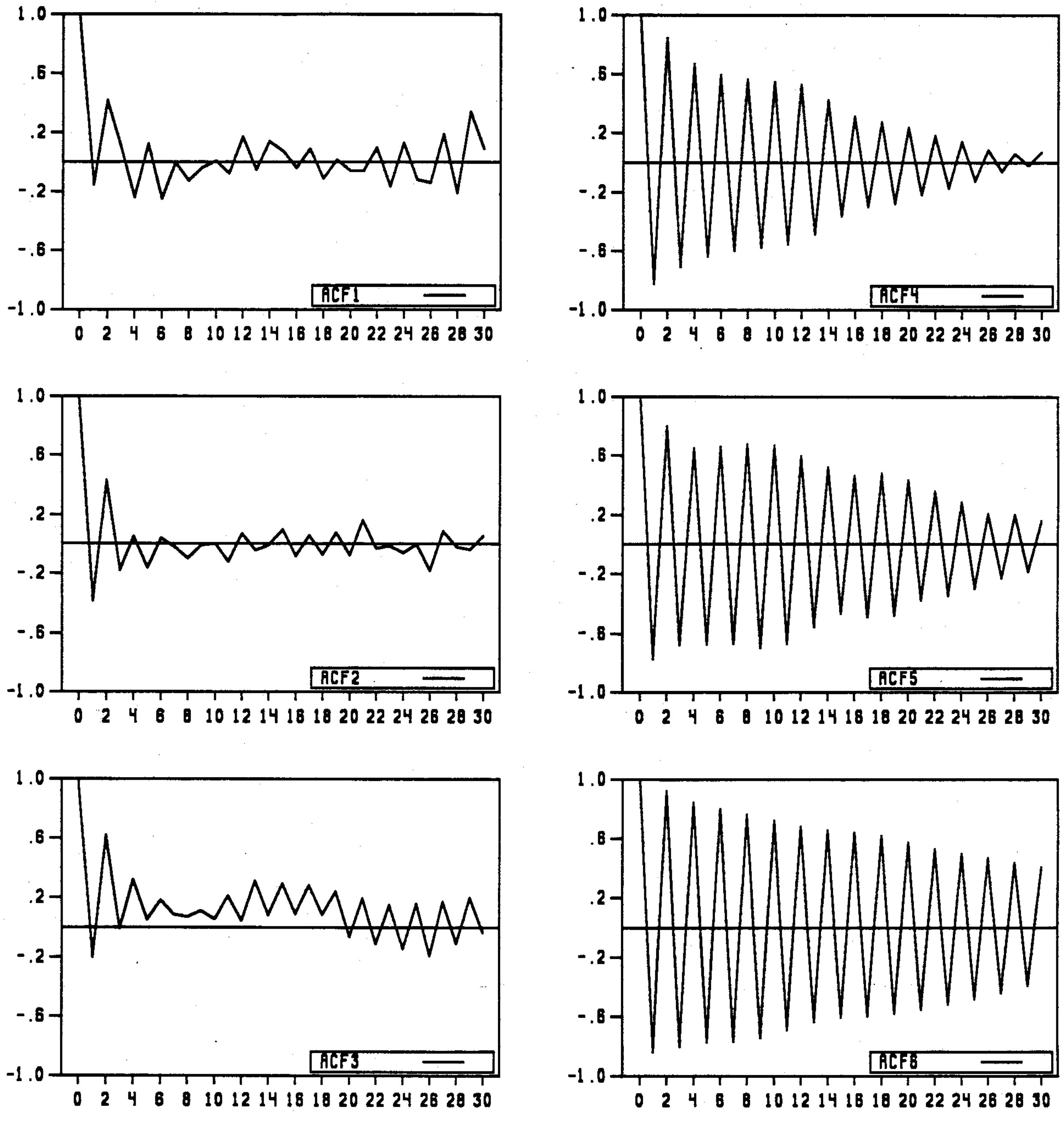

FIGURE 7c: Autocorrelation functions of the six components of the Finnish system which emerge as solutions of the seasonal cointegration problem at frequency π. All components filtered by $1-L+L^2-L^3$

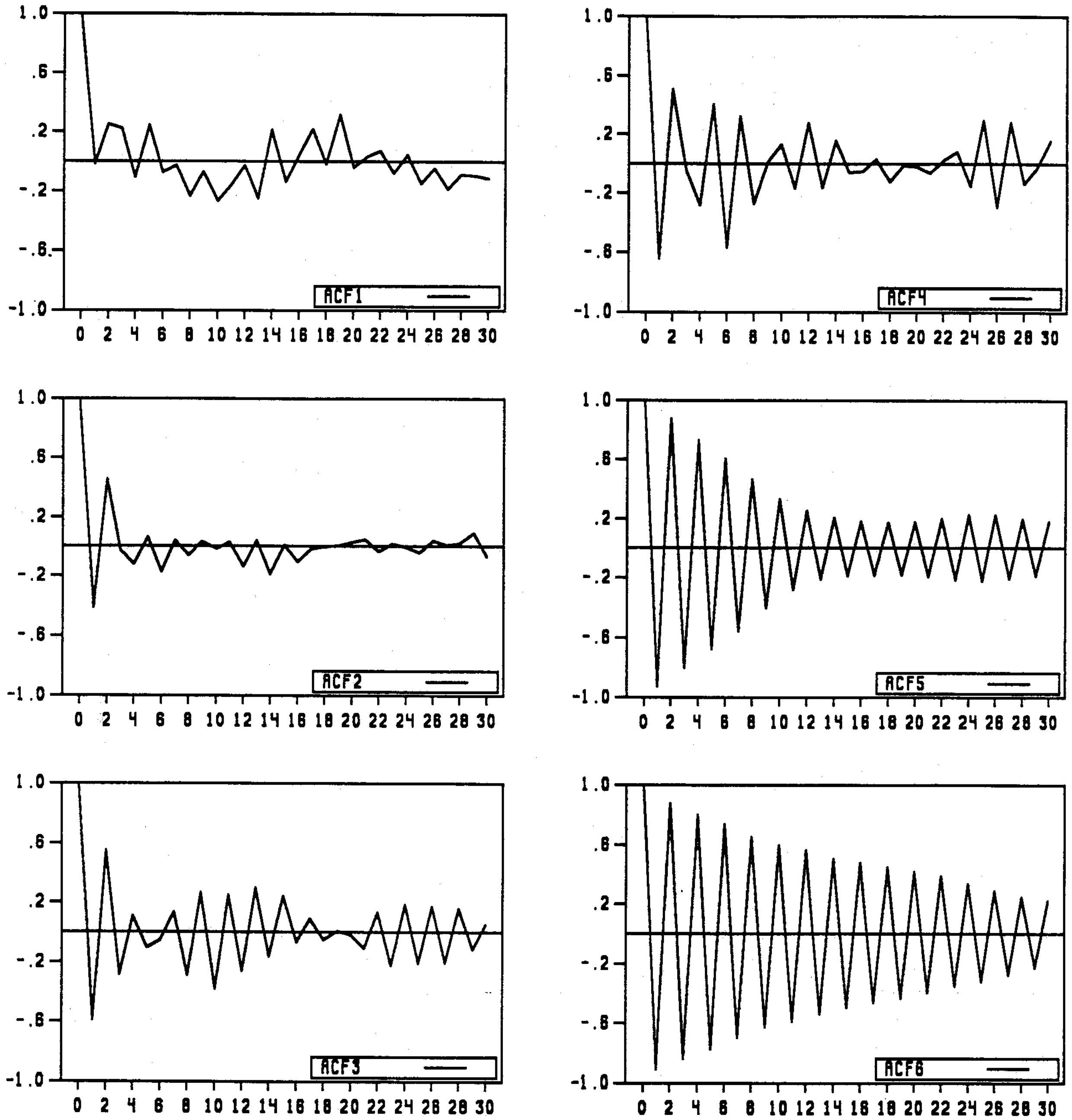

FIGURE 7d: Autocorrelation functions of the six components of the British system which emerge as solutions of the seasonal cointegration problem at frequency π. All components filtered by $1-L+L^2-L^3$

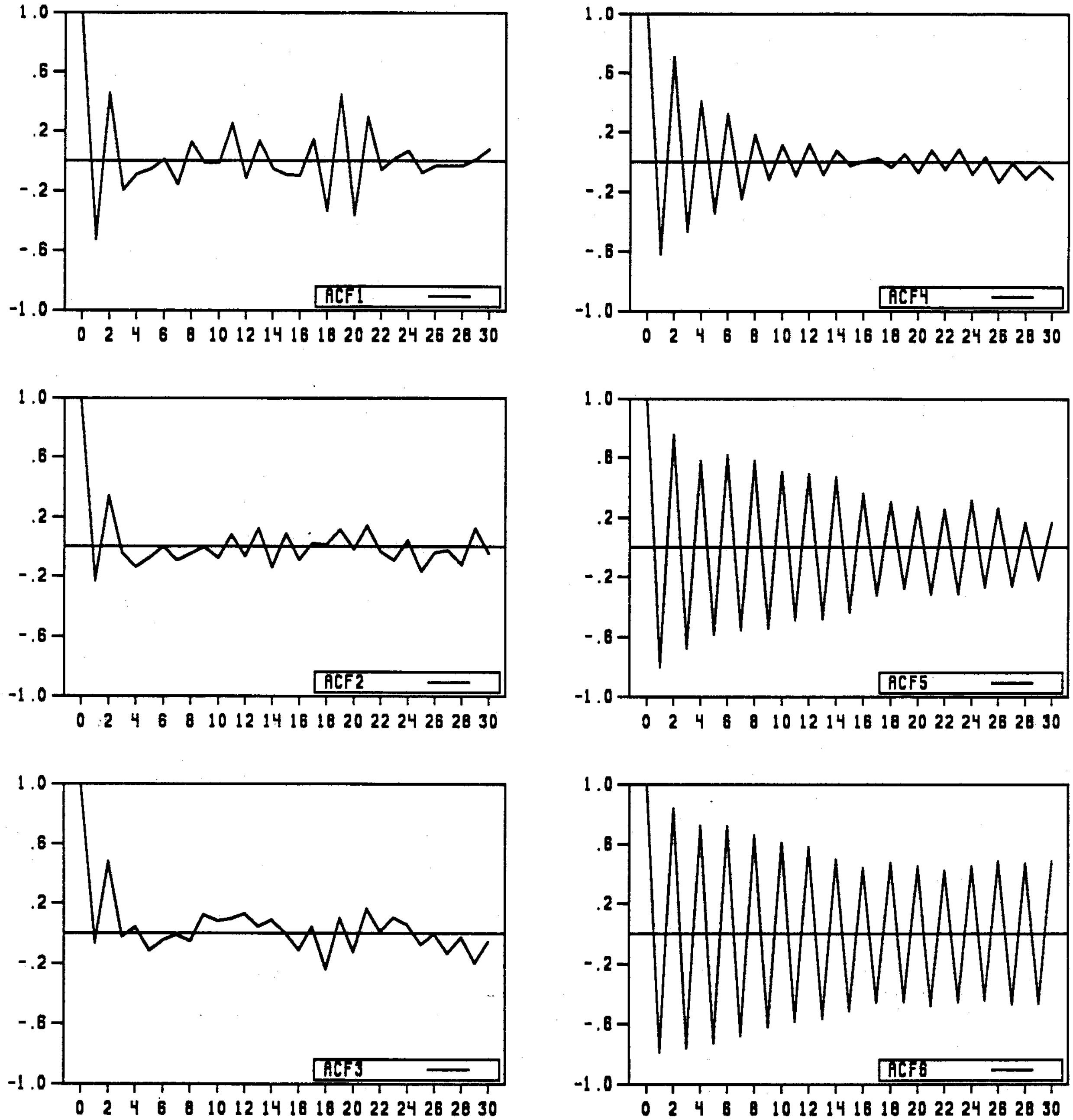

FIGURE 8a: Autocorrelation functions of the six components of the Austrian system which emerge as solutions of the seasonal cointegration problem at frequency $\pi/2$. All components filtered by $1-L^2$

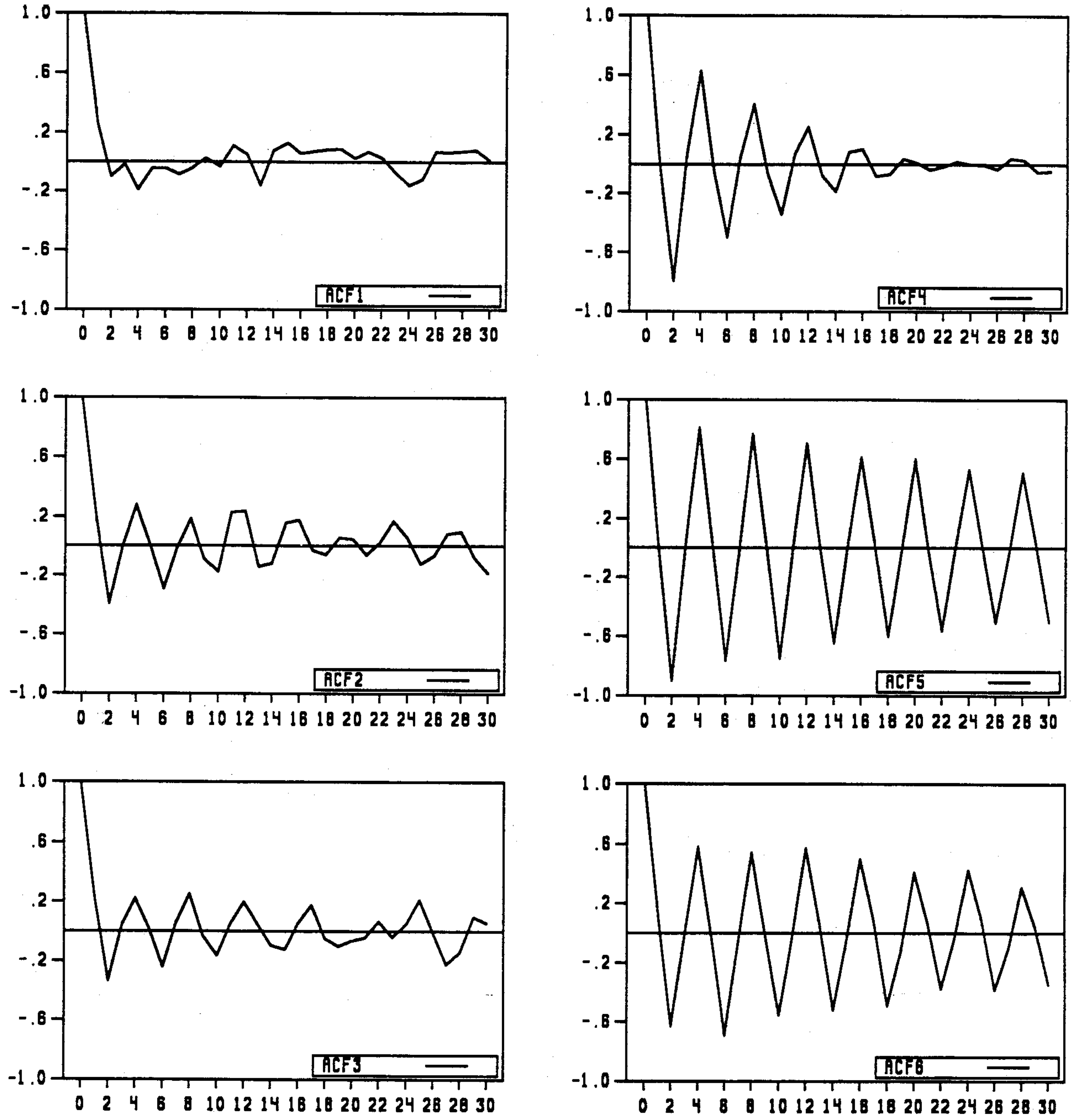

FIGURE 8b: Autocorrelation functions of the six components of the German system which emerge as solutions of the seasonal cointegration problem at frequency $\pi/2$. All components filtered by $1-L^2$

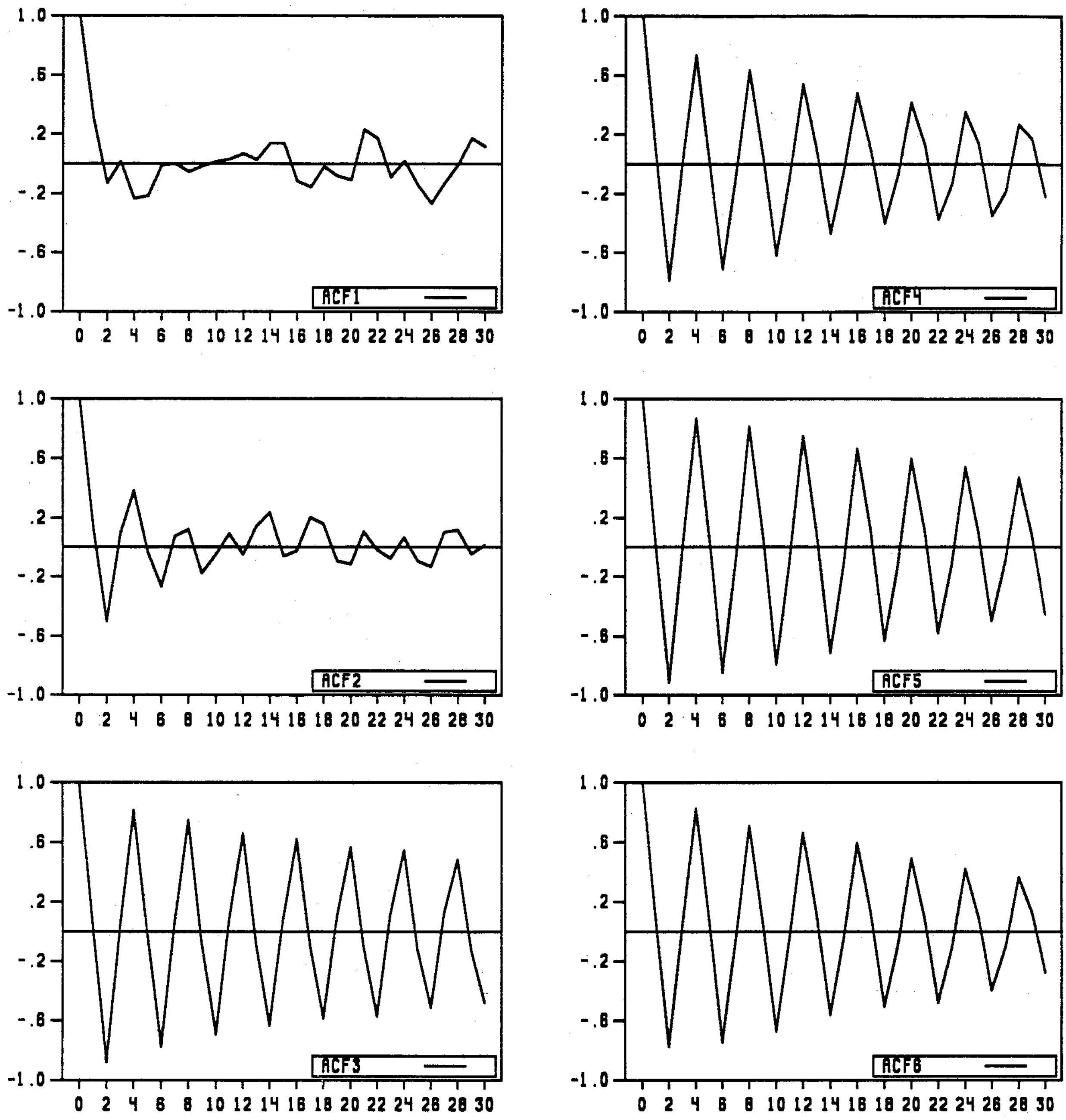

FIGURE 8c: Autocorrelation functions of the six components of the Finnish system which emerge as solutions of the seasonal cointegration problem at frequency $\pi/2$. All components filtered by $1-L^2$

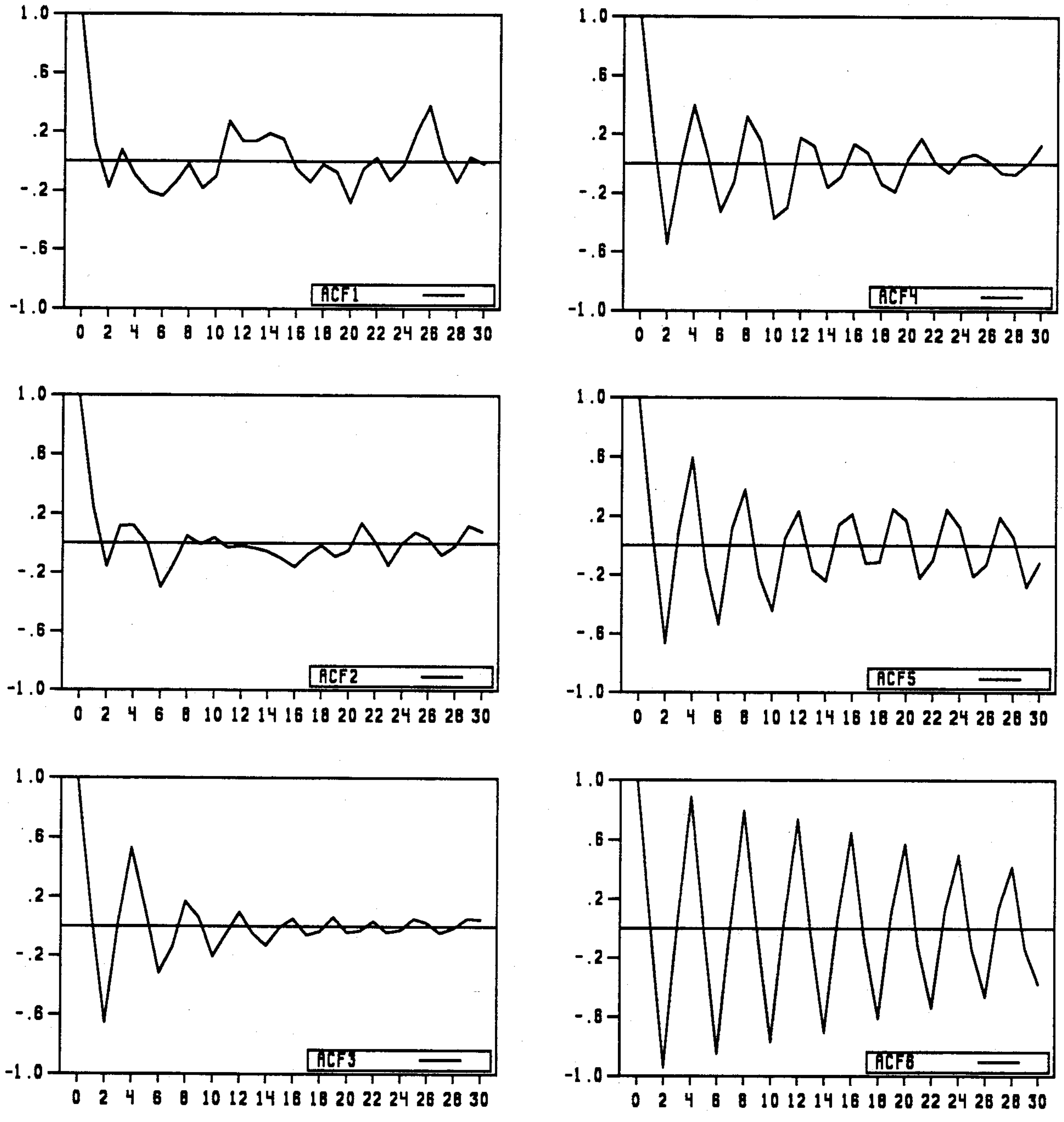

FIGURE 8d: Autocorrelation functions of the six components of the British system which emerge as solutions of the seasonal cointegration problem at frequency $\pi/2$. All components filtered by $1-L^2$