Biased Estimation in a Simple Extension of a Standard Error Correction Model*

Christian Müller

1. Introduction

The concept of cointegration (see e.g., Engle and Granger, 1987; Johansen, 1988) has been extensively used to model economic equilibrium relationships (see e.g., Juselius, 1990; Johansen, 1995; Hubrich, 2001). The links between economic and econometric concepts in modelling equilibria are now well understood and are part of the standard tools of empirical analysis. Loosely speaking, economic equilibrium relationships have their counterparts in cointegration, or, more generally speaking, error correction relationships whose existence can be tested and whose parameters can be estimated. The other side of the coin is the necessary adjustment back to the equilibrium once it has temporarily been distorted. This adjustment mechanism has also been analysed quite extensively. For instance, Ericsson, Hendry, and Mizon (1998) and Ericsson (1992) look at the implications for inference in cointegrated systems in the presence, or rather absence of equilibrium adjustment in one direction or another.

The reactions to deviations from equilibrium have also been interpreted as evidence for causality or non-casuality of variables within a system. Applying Hosoya and Yuzo's (1991) strength of causality measure, Granger and Lin (1995) show that for nearly integrated systems lack of adjustment to equilibrium of one variable can be considered evidence for long run Granger causality of that variable for the other one in a bivariate system. The relationship between error correction and causality has also become popular in applied research. Among others Juselius (1996), Lütkepohl and Wolters (2003) and Juselius and

MacDonald (2004) use it to qualify certain variables as causal for other variables based on the characteristics of the equilibrium correction mechanism.

Looking at vector autoregressive (VAR) models for policy analysis and advice, Hendry and Mizon (1998) point out that potential policy instruments must not feature error correction behaviour if they are supposed to have a lasting impact. Finally, Pesaran, Shin, and Smith (2000) argue that knowledge about the directions of error correction yields useful information about large scale VAR modelling which would otherwise be haunted by the curse of dimensionality.

This paper focuses on the determination of the adjustment mechanism in equilibrium or error correction models. It develops an example of an economic model which is fairly general in nature and which can easily be estimated with standard econometric tools. It will be shown that under moderate assumptions these tools are producing misleading results, i.e. the equilibrium adjustment coefficients are estimated with a bias which does not disappear asymptotically. This effect is obtained by considering expectations in the data generating process. In the light of the above mentioned references this can be considered rather bad news since the interpretation of the econometric results would be severely misguided.

The remainder of the paper is organised as follows. After a brief description of the problem, empirical examples in Section 3 illustrate the main issues and Section 4 discusses the results. An informal test is proposed to cope with the issue while a rough formal statement of the problem is sketched in the appendix.

2. An Economic Model and the Econometric Approach to It

2.1 The Model

We consider a variable or set of variables y which is a function of another variable or set of variables z. The realisations of y and z through time are denoted y_t and z_t respectively. Furthermore, expectations at time t about a value of a variable x_t at time $t+s$ is denoted $E_t(x_{t+s})$. The relationship between y_t and z_t is given by

$$y_t = \Lambda E_t(z_{t+s}) + \eta_{1,t}, \, s > 0,$$
$$z_t = \Phi z_{t-1} + \eta_{2,t},$$
$$\eta_{i,t} = A_i(L)\varepsilon_{i,t}, \, i = 1, 2,$$
$$E_t(z_{t+s}) = z_{t+s} + \iota_{t+s}, \, s \in Z,$$
$$\text{and } E_t(z_{t+s} \mid z_{t+s-j}) \neq E_t(z_{t+s}), \text{ for some } j \geq s. \tag{1}$$

As usual, L and Δ are the lag and first difference operator respectively with $x_t L^i = x_{t-i}$ and $\Delta = 1 - L$. The terms $A_i(L) = I_{n_i} - A_{i,1}L - A_{i,2}L^2 - \cdots - A_{i,p+1}L^{p+1}$, $i = 1,2$, denote polynomials in the lag operator of length $p + 1$ at most. Their roots are strictly outside the unit circle and the innovations $\varepsilon_{i,t}$, $i = 1,2$ are independent multivariate white noise. The $(n_2 \times n_1)$ matrix Λ' has full column rank n_1 with $1 \leq n_1 \leq n_2$, $n_1, n_2 \in N$.

As I am interested in reasonable forecasts only, it is further assumed that ι_t – the forecast error – is a stationary variable. With ι_t stationary the expected forecast error variance does not increase over time. As a side effect the long–run properties of the system are exclusively ruled by Φ only.

A situation such as in Equation (1) very often arises in macroeconomic theory. For example, the real interest rate (playing the role of y_t) is very often considered nothing but the sum of the real interest rate plus expected future inflation (this sum being z_{t+s} here). Technically, it is easy to see that (1) is a simple augmentation of the process considered by ENGLE and GRANGER (1987) whose system had $p = s = 0$, $\iota = 0$ $\forall t$, and $\Phi = I_{n_2}$.

The following comments seem to be in order. First, the linear relationship between y_t and z_t shall be considered an approximation of the true but probably unknown underlying function. The error processes $\eta_{i,t}$, $i = 1,2$ are approximations of the (true) structure of the data generating process of y_t and z_t which is not captured by the assumed functional form. This interpretation can be justified by the fact that autoregressive processes in general represent useful linear approximations to a wide range of (possibly also nonlinear) functions. Finally, under rational expectations one could consider the additional assumption $E(l_{t+s}, z_{t+s-j}) = 0$ $\forall s > 0$, $j > 0$, which is not needed though for the results below.

Next, the model is transformed into an estimable version. For simplicity we first chose $s = 1$ and generalise the result later on. Define

$$
A_0 = \begin{bmatrix} I_{n_1} & -\Lambda & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & I_{n_2} & -\Phi \\ 0_{n_2 \times n_2} & 0_{n_2 \times n_2} & I_{n_2} \end{bmatrix} \quad A_1 = \begin{bmatrix} 0_{n_1 \times n_1} & 0_{n_1 \times n_2} & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & 0_{n_2 \times n_2} & 0_{n_2 \times n_2} \\ 0_{n_2 \times n_2} & 0_{n_2 \times n_2} & \Phi \end{bmatrix}
$$

$$
\eta_t = \begin{bmatrix} \eta_{1,t} + \Lambda \iota_{t+1} \\ \eta_{2,t+1} \\ \eta_{2,t} \end{bmatrix} \quad Y_t = \begin{bmatrix} y_t \\ z_{t+1} \\ z_t \end{bmatrix}
$$

and write (1) as

$$A_0 Y_t = A_1 Y_{t-1} + \eta_t, \tag{2}$$

where $0_{l_1 \times l_2}$ denotes a $(l_1 \times l_2)$ matrix of zeros and I_l a l-dimensional identity matrix. Pre-multiplying (2) by the inverse of A_0 obtains

$$
\begin{bmatrix} y_t \\ z_{t+1} \\ z_t \end{bmatrix} = \begin{bmatrix} 0_{n_1 \times n_1} & 0_{n_1 \times n_2} & \Lambda\Phi^2 \\ 0_{n_2 \times n_1} & 0_{n_2 \times n_2} & \Phi^2 \\ 0_{n_2 \times n_2} & 0_{n_2 \times n_2} & \Phi \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_t \\ z_{t-1} \end{bmatrix} + A_0^{-1}\eta_t ,
$$

$$= A_1^* Y_{t-1} + \eta_t^* \tag{3}$$

In the next step, (3) is transformed into the error correction form by subtracting Y_{t-1} from both sides to arrive at

$$
\triangle Y_t = \begin{bmatrix} -I_{n_1} & 0_{n_1 \times n_2} & \Lambda\Phi^2 (10) \\ 0_{n_2 \times n_1} & -I_{n_2} & \Phi^2 (11) \\ 0_{n_2 \times n_2} & 0_{n_2 \times n_2} & -(I_{n_2} - \Phi) \end{bmatrix} Y_{t-1} + \eta_t^*
$$

$$= \Pi Y_{t-1} + \eta_t^* . \tag{4}$$

The matrix Π can be decomposed into two full column rank matrices α and β with

$$
\alpha = \begin{bmatrix} -I_{n_1} & 0_{n_1 \times n_2} & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & -I_{n_2} & 0_{n_2 \times n_2} \\ 0_{n_2 \times n_2} & 0_{n_2 \times n_2} & -I_{n_2} \end{bmatrix} \quad \beta' = \begin{bmatrix} I_{n_1} & 0_{n_1 \times n_2} & -\Lambda^* \\ 0_{n_2 \times n_1} & I_{n_2} & -\Phi^2 \\ 0_{n_2 \times n_1} & 0_{n_2 \times n_2} & \Psi \end{bmatrix}
$$

where $\Psi = I_{n_2} - \Phi$, $\Lambda^* = \Lambda\Phi^2$, and $\Pi = \alpha\beta'$.[1] In general, the matrix Π is unknown and has to be estimated. Likewise, generally neither α nor β are identified. Given, however, the identification of β, α can be identified too. The

---

1   Note that for $\Phi = I_{n_2}$ (2) is a cointegrated system and the last columns of α and β would vanish.

identification of β is usually the first and more interesting concern as it can be achieved using restrictions arising from economic theory such as the Fisher relationship or interest parity conditions. In what follows the focus will be on the estimation of α conditioning on standard approaches for identifying β. It will be demonstrated that some standard econometric technique to handle models like (1) by (3) or rather, some derivative of (3) turns out inappropriate.

2.2 The Estimation Approach

The econometric counterpart of (1) can be and usually is set up as an error correction system. In case of $\Phi = I_{n_2}$ (1) is a cointegrated system with $\beta' = (I_{n_1} : -\Lambda)$ as the cointegration matrix its rank being n_1. The corresponding estimation model can be written as

$$\Delta Y_t = \alpha \beta' Y_{t-1} + \sum_{i=1}^{d} \Gamma_i \Delta Y_{t-i} + \eta_t \tag{5}$$

with $Y_t = (y'_t, z'_t)'$, the $(n \times n_1)$ coefficient matrices α and β, the $(n \times n)$ coefficient matrices Γ_i with $n = n_1 + n_2$ and $p \leq d$, and the $(n \times 1)$ vector of innovations $\eta_t = (\eta'_{1,t} + \eta'_{2,t}\Lambda' + \iota'_{t+s}\Lambda', \eta'_{2,t})'.^2$ In the present context, n_1 corresponds to the number of cointegrating relationships and n_2 is the number of long-run driving forces, or the number of stochastic trends in the system. The sum $n = n_1 + n_2$ matches the number of variables in Y_t. Putting aside the estimation of the Γ_i we focus on the derivation of the estimates for α. As GRANGER and LIN (1995) have shown the structure of α can be used to make statements about the subset of Y_t that is driving another subset of Y_t in the long run. Note that inference is usually only made with respect to z_t but not with respect to z_{t+s} even though (3) might imply so. The interpretation of z_t being Granger causal for y_t is nevertheless true as is the notion that z_{t+s} is Granger causal for y_t.

Letting α_1 correspond to y_t and α_2 to z_t conditions 1 and 2 represent necessary (but not sufficient) conditions for model (1) to be congruent with the actual observations.

Condition 1: $\alpha_2 = 0_{n_2 \times n_1}$
Condition 2: $\alpha_1 \neq 0_{n_1 \times n_1}$

Estimating (5) should thus yield estimates for α which comply with conditions 1 and 2. It is noteworthy that the elements of α_1 should in general be negative in

---

2   The lagged dependent variables in (5) are meant to account for the autocorrelation structure in $\eta_{i,t}$ which is therefore ignored.

order to obtain a stationary system. Therefore, condition 2 is rather weak. The appendix shows that under pretty reasonable circumstances a clear statement about the estimation outcome cannot be made even if the data is generated in line with (1). In general, the estimates for α_1 and α_2 will both be biased upward. Two main factors drive the outcome.

Let $\hat{\alpha}_i, i = 1,2$ denote the ordinary least squares estimator for $\alpha_i, i = 1,2$ when Λ is known, $a_i, i = 1,2$ be the probability limit of $(\alpha_i - \hat{\alpha}_i)$, $i = 1,2$. Furthermore, the positive semi-definite Σ_{ι_t} and the positive definite $\Sigma_{\eta_{i,t}}$, $i = 1,2$ matrices be the variance–covariance matrices of ι_t and $\eta_{i,t}$, $i = 1,2$, respectively. Assuming that model (1) holds, it can be shown that

1. *Good-forecast-bias:* If there exist forecast errors ι_t^* and ι_t^+ with corresponding $a_2(\iota_t^*)$, $a_2(\iota_t^+)$, and $\Sigma^{d\iota} = \Sigma_{\iota_t^+} - \Sigma_{\iota_t^*}, \Sigma^{d\iota}$ being positive semi-definite, $a_2(\iota_t^*) > a_2(\iota_t^+)$.
2. *No-poor-model-bias:* For any innovation processes $\eta_{1,t}^*$ and $\eta_{1,t}^+$ related to y_t in (1) with corresponding $a_1(\eta_t^*)$, $a_1(\eta_t^+)$, and $\Sigma^{d\eta} = \Sigma_{\eta_t^+} - \Sigma_{\eta_t^*}$, $\Sigma^{d\iota}$ being positive semi-definite, $a_1(\eta_t^*) < a_1(\eta_t^+)$.

The above list implies that the results do not improve if, for example, expectations approach perfection ($\iota_t \to 0_{n_2} \; \forall t$). Instead, the estimates do not improve but worsen. The same holds if the variance in $\eta_{1,t}$ becomes very large. As an extreme case, one could even obtain estimates for α where α_1 meets the condition 1 and α_2 meets condition 2.

This of course is rather bad news. A logical implication is that the presented and very widely used empirical approach cannot discriminate between having made a mistake when building the model or not. Section 4 proposes a convenient albeit not always feasible procedure to identify and circumvent this pitfall.

3. Empirical Examples

The Fisher relation (Fisher, 1930) and the uncovered interest rate parity condition are popular examples of economic hypotheses with rational expectations, among others. In all these cases the objective is to explain the levels of (typically) the long-term interest rate. To make the discussion below more handy this paper uses the terminology of Granger and Lin (1995) where the weakly exogenous variables are referred to as causal in the long–run.

3.1 The Fisher Relation

The starting point is the notion that rational individuals focus on the real return on investments, that is, after accounting for inflation. Therefore, the (long-term) nominal interest rate (i_t^l) needed to convince people to lend money is the sum of the desired real return (real interest rate, r_t) plus the expected inflation $(E_t(\pi_{t+s}))$:

$$i_t^l = r_t + E_t(\pi_{t+s}) \tag{6}$$

The difficulty that arises, is, of course, that expected inflation is not observable. That's why in empirical work it is often approximated by the current inflation rate, which would be the best linear forecast if inflation followed a random walk, for example.

Therefore, the Fisher hypothesis can be cast in the framework of Section 2 with i_t^l being the endogenous variable and π_t playing the role of z_t.

3.2 The Uncovered Interest Rate Hypothesis

Taking again the perspective of an investor, the portfolio choice will also be made considering foreign bonds. If the foreign bond rates are determined exogenously (e.g., the U.S. bonds with respect to the rest of the world), then the choice to buy or sell domestic bonds will depend on what is expected about the future level of the foreign alternative. Again, the role of expectations becomes central and the setting of Section 2 applicable.

In all these examples autoregressive processes are commonly used to establish a link between the observable values of the exogenous variable and the unobservable expectations about it.

3.3 Estimation Results

The following exercise presents results for the USA, Germany and Switzerland. The standard setup is a reduced rank regression as it has been suggested by JOHANSEN (1988). In all cases, the choice of variables makes sure that the cointegration rank, as implied by the theory, is exactly one. The test statistic for the cointegration rank test is also provided. The general model for estimation is (5). The lag order p is chosen according to selection criteria. If the suggested lag order is not sufficient to account for residual autocorrelation further lags are added. Most of the time, this procedure solves the problem. In one instance

(example 1 below), the residual autocorrelation cannot be coped with in the multivariate setting. Therefore, single equation methods are also used. With these, a more flexible lag structure can be implemented that also solves the problem of autocorrelation.

Another difficulty with the data is heteroscedasticity and non-normality of residuals which can often be observed when modelling interest rates. Here, no definite answer can be given. It has not always been possible to eliminate *ARCH* effects and excess kurtosis. All results are presented in Table 1 except for the residual properties which are, of course, available on request.

In Table 1, the information regarding the model setup is in columns 1–7. In all cases the cointegration rank test supports the hypothetical number of cointegration relations. In the column labelled "β coefficients", it is checked whether the hypothetical cointegration coefficients can be imposed. These coefficients also imply that z_t is an unbiased estimate for $E_t(z_{t+1})$. Again, this is the case in almost all instances at the 10 percent level of significance. Where this is the case (examples 1–3), the following test for the restrictions on the adjustment coefficients (α) is performed including the restrictions on the β coefficients.

Each of the first lines in the examples 1–6 should, according to the outlines above, feature a rejection of the null hypothesis that the respective adjustment coefficient is zero. It should be born in mind that this variable is always supposed to represent the independent variable for the long-run relationship in economic terms. As expected, the estimation results seem to produce the opposite conclusion, namely that the presumed causal variable significantly reacts to deviations from equilibrium while the theoretically endogenous variable (2nd line) does not.

For example, in case 1 where the Fisher parity is tested for U.S. data, the hypothesis is that inflation expectations rule the nominal interest rates. In the econometric model, the expectations are replaced by current inflation which is viewed as a predictor of unobservable inflation expectations. Obviously and independent of the specific model, the null hypothesis that inflation does not adjust to deviations from the long-run equilibrium is strongly rejected. At the same time, however, it is found that interest rates do not adjust significantly. While the latter statement was found to be true at the 10 percent level only, the situation is much clearer in the Swiss case (example 2). Here, the hypothesis that interest rates do not adjust cannot be rejected at the 18 percent level.

Example 3 is concerned with the interest rate parity between Germany and Switzerland. From the Swiss perspective, Germany is a large economy whose bond rates appear exogenous with respect to the Swiss rates. Therefore, the Swiss National Bank would be forced to keep an eye on the German rate if too strong

Table 1: Empirical Evidence

No.	Variables[a]	Country	Sample	H_0^0	LR[b]	q	H_0^1	LR stat.[prob.]	H_0^2	LR stat.[c] [prob.]	Method[d]
				Cointegration Test			β coefficients		α coefficients		
Fisher Relation											
1	CPI infl. (β_1)	USA	$89{:}12 - 03{:}08$	$rk = 0$	22.76 [.02]	5	$\beta_1 = -\beta_2 = 1$	$\chi^2(1) = 3.15$ [.07]	$\alpha_1 = 0$	$\chi^2(1) = \mathbf{13.06}$ [.00]	MV
	Bond y. (β_2)	USA	$T = 161$	$rk = 1$	3.28 [.54]				$\alpha_2 = 0$	$\chi^2(1) = 2.70$ [.10]	
				$rk = 0$	74.67 [.00]	2	$\beta_1 = -\beta_2 = 1$	$\chi^2(1) = 13.98$ [.00]	$\alpha_1 = 0$	$\chi^2(1) = \mathbf{64.38}$ [.00]	MV
				$rk = 1$	3.45 [.51]				$\alpha_2 = 0$	$\chi^2(1) = 2.70$ [.10]	
									$\alpha_1 = 0$	$\chi^2(1) = \mathbf{9.5}$ [.00]	SEQ
									$\alpha_2 = 0$	$\chi^2(1) = 1.94$ [.16]	
2	CPI infl. (β_1)	CH	$90{:}02 - 03{:}07$	$rk = 0$	20.50 [.05]	4	$\beta_1 = -\beta_2 = 1$	$\chi^2(1) = 1.78$ [.18]	$H_0^1 \wedge \alpha_1 = 0$	$\chi^2(2) = \mathbf{15.86}$ [.00]	MV
	Bond y. (β_2)	CH	$T = 158$	$rk = 1$	2.61 [.66]				$H_0^1 \wedge \alpha_2 = 0$	$\chi^2(2) = 3.34$ [.18]	
Uncovered Interest Rate Parity Condition											
3	LIBOR (β_1)	GER	$92{:}01 - 03{:}07$	$rk = 0$	744.41 [.00]	2	$\beta_1 = -\beta_2 = 1$	$\chi^2(1) = 2.73$ [.10]	$H_0^1 \wedge \alpha_1 = 0$	$\chi^2(2) = \mathbf{13.26}$ [.00]	MV
	LIBOR (β_2)	CH	$T = 139$	$rk = 1$	6.23 [.18]				$H_0^1 \wedge \alpha_2 = 0$	$\chi^2(2) = 2.80$ [.25]	

a   CPI denotes consumer price index, Bond y. is short for government bond yield, LIBOR is the interest rate for short term credits (3-months) at the London interbank market, and Money is the interest rate on one-month interbank credits. More details can be found in Table 3.

b   Likelihood ratio test for the cointegration rank test [Johansen, 1995, tab. 15.2].

c   One degree of freedom if no restriction on β-vector imposed, 2 degrees of freedom if H_0^1 is also imposed (no. 2,3,5–6), 6 degrees of freedom (no. 4): H_0^1 and $\alpha_1 = \alpha_3 = 0$ additionally imposed

d   MV abbreviates multivariate model, SEQ single equation model.

a revaluation of the Swiss Franc versus the Euro (or Deutschmark) is considered not desirable. The way to ensure this most efficiently is, of course, to anticipate future movements of the German rate. Consequently, even though the German rate is the long-run driving force with respect to the Swiss rate, according to the model, the adjustment coefficients should seemingly imply the opposite. This is actually the case. The hypothesis that German rates do not adjust to Swiss rates is very strongly rejected while the hypothesis that Swiss rates do not adjust to German rates passes the test.

4. Discussion

4.1 Is There a Cure?

Having described and illustrated the problem, a natural question is of course whether there is a cure for it. The most desirable remedy would be an estimation setup where the economic model itself can be tested directly. In the standard situation, an indirect approach is used because the key element, the expectation about z_t, is not observable. Replacing it by an (unbiased) estimator helps to circumvent the measurement problem yet incurs the paradox. This point can be illustrated by the following additional regression, where the UIP between German and Swiss interest rates is used again. This time however, the unobservable expected German rate is approximated by a very good predictor, which is its own future realisation. Table 2 has the details.

Obviously, the three months ahead realisation of the German 3-months interest rate is a good guess about the German 3-months interest rates three months ahead. A shock to this expectation (now) significantly affects the Swiss interest rate while no effect can be measured in the opposite direction. Thus, the paradox is solved "econometrically". In regression 4 of Table 2, the economic and econometric notion of dependence and independence finally coincide.[3] A crosscheck is provided by example 5, where instead of the German rate, the Swiss rate is leading three periods, the outcome however, is qualitatively the same as that of model 3.

---

3   Note that theoretically, lagging one variable of the system should not alter the cointegration test results. In the empirical example it does so. However, this alterations does not matter because in a stationary system – as it is implied by the tests in models number 5 and 6 – the framework of Section 2 in principal still holds without the additional complication of non-stationaritites.

Table 2: Empirical Evidence 2nd Step

No.	Variables[a]	Country	Sample	Cointegration Test			β coefficients		α coefficients			Method[d]
				H_0^0	LR[b]	q	H_0^1	LR stat.[prob.]	H_0^2	LR stat.[c] [prob.]		
Uncovered Interest Rate Parity Condition												
4	$LIBOR_{t+3}$ (β_1)	GER	$92:01-03:04$	$rk=0$	25.24 [.00]	3	$\beta_1=-\beta_2=1$	$\chi^2(1)=.63$ [.43]	$H_0^1 \wedge \alpha_1=0$	$\chi^2(2)=.72$ [.70]		MV
	LIBOR (β_2)	CH	$T=136$	$rk=1$	11.42 [.02]				$H_0^1 \wedge \alpha_2=0$	$\chi^2(2)=\mathbf{9.39}$ [.01]		
5	LIBOR (β_1)	GER	$92:01-03:04$	$rk=0$	60.727 [.00]	3	$\beta_1=-\beta_2=1$	$\chi^2(1)=2.56$ [.01]	$H_0^1 \wedge \alpha_1=0$	$\chi^2(2)=\mathbf{39.97}$ [.00]		MV
	$LIBOR_{t+3}$ (β_2)	CH	$T=136$	$rk=1$	15.0 [.00]				$H_0^1 \wedge \alpha_2=0$	$\chi^2(2)=2.58$ [.27]		
6	$Money_{t+3}$ (β_1)	CH	$89:05-03:04$	$rk=0$	50.089 [.00]	4	$\beta_1=-\beta_2=1$	$\chi^2(1)=3.82$ [.05]	$\alpha_1=0$	$\chi^2(2)=1.70$ [.19]		MV
	LIBOR (β_2)	CH	$T=168$	$rk=1$	2.64 [.66]				$\alpha_2=0$	$\chi^2(2)=\mathbf{49.58}$ [.00]		

a   CPI denotes consumer price index, Bond y. is short for government bond yield, LIBOR is the interest rate for short term credits (3-months) at the London interbank market, and Money is the interest rate on one-month interbank credits. More details can be found in Table 3.

b   Likelihood ratio test for the cointegration rank test [JOHANSEN, 1995, tab. 15.2].

c   One degree of freedom if no restriction on β-vector imposed, 2 degrees of freedom if H_0^1 is also imposed (no. 2,3,5–6), 6 degrees of freedom (no. 4): H_0^1 and $\alpha_1=\alpha_3=0$ additionally imposed

d   MV abbreviates multivariate model, SEQ single equation model.

Unfortunately, there are not always good predictors at hand. For example, when testing the Fisher parity for long-term bonds, it is not clear how the future inflation rates should be weighted in order to produce a good proxy for the inflation in the remaining time to maturity. Similar arguments hold for many other relationships.

4.2 Relevance

The literature has so far paid not too much attention to the seemingly surprising lack of weak exogeneity of the supposed long-run driving variables. There are, however, also good reasons for that. For example, it is of interest in itself if the spread between nominal interest rates and inflation is stationary or not, because it helps to learn about the Fisher hypothesis. The same holds for the other concepts briefly discussed. This inference can be made without reference to the adjustment characteristics as long as there is adjustment towards equilibrium at all.

Another stream of literature makes use of the fact that y_t needs to be a good *predictor* for *future* z_{t+s} if (1) is the true model. Thus, regressing z_t on y_{t-s} (or, rather on $(z_{t-s} - y_{t-s})$ should yield a significant coefficient and significance would be interpreted being consistent with the economic model. This conclusion, however clouds the fact that according to (1) variations in z_{t-s} would have to affect y_t highlighting that the coefficients of such a model are very likely inefficiently estimated and more or less useless for economic policy analysis (see Ericson et al., 1998, p. 377). The latter is the ultimate goal of many econometric studies, however.
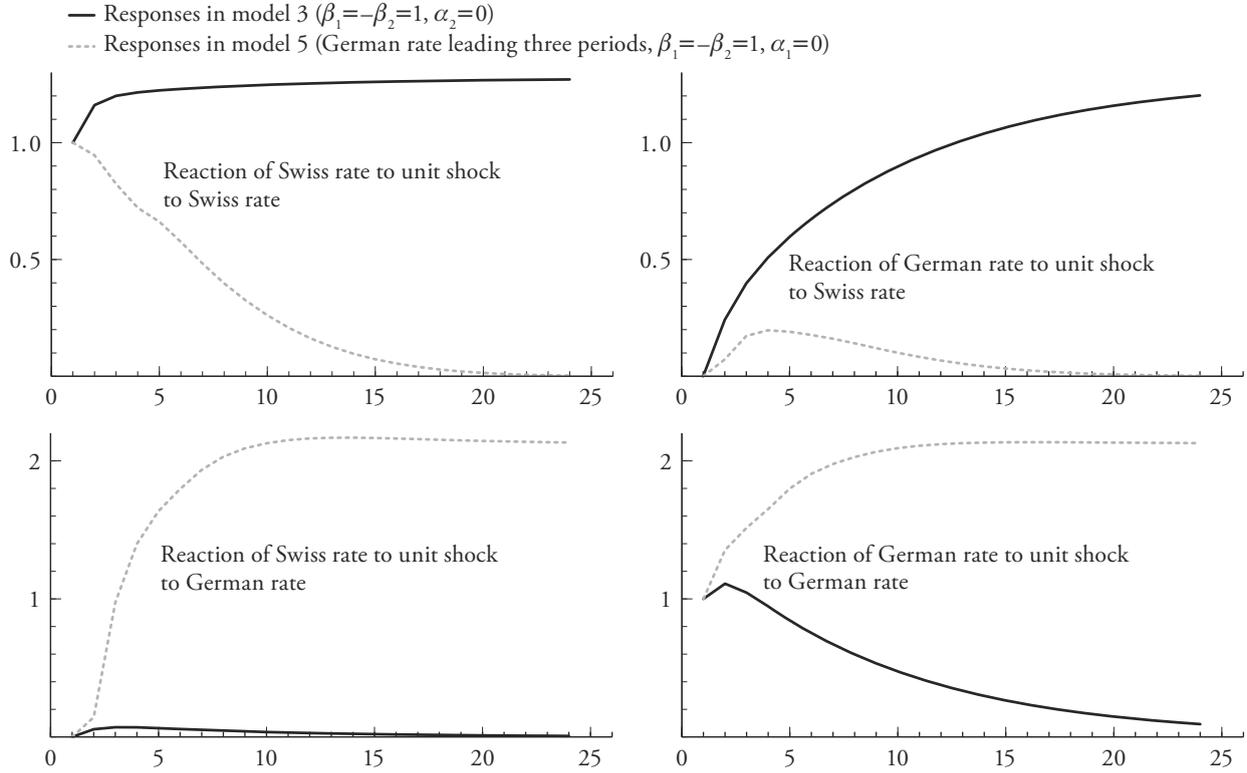
4.2.1 Forecasting and Policy Simulation

Following up on the last point, there are also at least two situations where the difference matters. The first is forecasting.[4] Figure 1 illustrates the effect. Systems 3 and 4 have been estimated and subjected to standard goodness-of-fit tests such as checking for any remaining autocorrelation in the residuals.[5] Based on these estimated models forecast error shocks can be simulated. This means that one equation is shocked once while no shock is allowed in the other equation at the same time. This resembles a hypothetical attempt of a policy maker who may regard either of the variables as a policy instrument. The corresponding reactions of the variables are then graphed.

---

4   Policy simulation and analysis are naturally related concepts, see Hendry and Mizon (1998), for example.
5   The test results indicate a good quality of the model such that it can be regarded data congruent.

Figure 1: Impulse-Responses in Systems 3 and 4.

Responses in model 3 ($\beta_1 = -\beta_2 = 1$, $\alpha_2 = 0$)
Responses in model 5 (German rate leading three periods, $\beta_1 = -\beta_2 = 1$, $\alpha_1 = 0$)

Reaction of Swiss rate to unit shock to Swiss rate

Reaction of German rate to unit shock to Swiss rate

Reaction of Swiss rate to unit shock to German rate

Reaction of German rate to unit shock to German rate

Obviously, the responses could hardly be more different. In model 3 the reaction of the Swiss rate to a shock in the German rate (lower left panel) dies out pretty quickly while in model 4 it remains above two for the whole simulation period.[6] Likewise striking is the reaction of the German rate in model 3 when the Swiss rate is shocked (upper right panel). It appears that the German rate responds strongly, while this cannot be observed in model 4.

Therefore, if one bases forecasts for the Swiss interest rate assuming a change in the ECB interest rate, for example, on model 3, not only would one obtain results which are at odds with conventional wisdom about the relationship between the German and Swiss economies, but one would also be diverted from the "true" causal links. Considering model 4 instead, solves the puzzle. These opposite reactions are a direct result of HENDRY and MIZON's (1998) analysis of instrument–target relationships. They show that weak exogeneity of the instrument variable (z in our case) with respect to Λ is a sufficient condition for a long–run zero response to a shock to the target variable (y in our case). The following argument shows that the choice between model 3 and 4 may not need to be purely arbitrary.

4.2.2 A Two-Stage-Procedure

A second situation where it may pay to account for the paradox is to test for the existence of the paradox itself. For example, for monetary policy analysis it would matter if money is causal for inflation or not. If one would assume that the demand for money were a function of the expected future price level instead of current prices an analysis based on impulse-responses can be severely impaired. In such cases rivalling economic models exist which pose causality in opposite directions and hence, not taking into account the possibility of biased estimates may result in a wrong conclusion. Even worse, since either direction of causality might be possible and since the $\hat{\alpha}$ would in both cases indicate equilibrium adjustment, the chance of ever noticing are very low.

In some situations, however, a not too difficult way exists to detect a bias. To see this, consider again the model of Section 2. If it was possible to replace the approximation of the expected value by the expectation itself, then the standard situation as of, e.g. ENGLE and GRANGER (1987), HENDRY and MIZON (1998), ERICSSON et al. (1998) arises. In terms of the stylised situation of Section 2, this

---

6   Note that no statement about significance with respect to the distance from zero is made. What matters most, however, is the (principal) difference between the responses in the two models.

results in estimates for α_1 and α_2 ($\hat{\alpha}_1$ and $\hat{\alpha}_2$, respectively) which are in accordance with the conditions 1 and 2.

The crucial point is that now $\hat{\alpha}_2$ will be zero if forecasts are nearly perfect. We now also obtain $\text{plim}(\hat{\alpha}_1) = \alpha_1$.[7] Thus, the economically sensible result is obtained which implies that z_t drives y_t in the long-run but not vice versa. Therefore, a two-step procedure can be proposed. First, the standard cointegration analysis is performed and the weak exogeneity properties are determined (see models 3). Then, the set of weakly exogenous variables, z_t, is replaced by its best possible s-step ahead forecast (which, e.g., could be z_{t+s}) and the analysis repeated (see models 4). If the results are identical to the ones obtained in the first step, one would be re-assured, that the underlying structural dependence, is as it appears to be from the face values of the estimates. If, however, some variables are now found to belong to the set y_t which in step 1 have been found belonging to z_t, then the true relationship is likely to be of the type sketched in Section 2.

Unfortunately, it is not always clear what the best possible forecast is. In the Fisher relationship, expected inflation is certainly not the inflation rate of a specific month in the future, but some "overall" future price change. That difficulty of course limits the potential for obtaining useful test results. The estimation result of models 3 versus 4 may represent examples where the two-step procedure proved useful, however.

5. Conclusion

In economic models where expectations about one variable rule the behaviour of another one the standard econometric approach is not very likely to reveal the true causal links if the expectations cannot be directly observed. This paper has shown that this result also holds for cointegrated relationships where the direction of adjustment towards the equilibrium is used to identify the long run dependent and independent variables. Moreover, a paradox may arise in which the true links are more likely recovered if the underlying economic model is in fact built on poor grounds. Therefore, when it comes to interpreting the adjustment coefficients, for example by impulse–response analysis one has to be particularly careful.

---

7   The expression plim denotes the probability limit. Of course, the value of α_1 is strictly speaking a function of $A_1(L)$ and $A_2(L)$. For the purpose of demonstration it is relevant to note that it will not be zero.

Various data examples using popular economic hypotheses have illustrated these considerations. The bias is especially relevant for forecasting and policy simulation. Under some circumstances, however, a simple cure for the paradox exists which also has the potential for testing for the true causal relations.

Appendix

A. From the Economic to the Econometric Approach

A.1 Step 1 – The Standard Regression

The model will first be re-stated and then the estimation of α will be discussed.

$$
\Delta \begin{bmatrix} y_t \\ z_{t+1} \\ z_t \end{bmatrix} = \begin{bmatrix} -I_{n_1} & 0_{n_1 \times n_2} & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & -I_{n_2} & 0_{n_2 \times n_2} \\ 0_{n_2 \times n_2} & 0_{n_2 \times n_2} & -I_{n_2} \end{bmatrix} \begin{bmatrix} I_{n_1} & 0_{n_1 \times n_2} & -\Lambda^* \\ 0_{n_2 \times n_1} & I_{n_2} & -\Phi^2 \\ 0_{n_2 \times n_1} & 0_{n_2 \times n_2} & \Psi \end{bmatrix}
$$

$$
\times \begin{bmatrix} y_{t-1} \\ z_t \\ z_{t-1} \end{bmatrix} + \eta_t^*
$$

In the standard estimation approach the second line is usually disregarded and inference is only made with respect to the first and last line of α as well as the first line of β'. For simplicity it is assumed that the matrix β can be super consistently estimated (e.g. in cointegrated systems), or that the economic prior regarding Λ is so strong that it need not be estimated at all. Then, the central casuality analysis is with respect to α alone. Let's write

$$
\Delta \begin{bmatrix} y_t \\ z_{t+1} \\ z_t \end{bmatrix} = \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \alpha_{3,2} \\ \alpha_{2,1} & \alpha_{2,2} & \alpha_{3,2} \\ \alpha_{3,1} & \alpha_{2,3} & \alpha_{3,3} \end{bmatrix} \begin{bmatrix} I_{n_1} & 0_{n_1 \times n_2} & -\Lambda^* \\ 0_{n_2 \times n_1} & I_{n_2} & -\Phi^2 \\ 0_{n_2 \times n_1} & 0_{n_2 \times n_2} & \Psi \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_t \\ z_{t-1} \end{bmatrix} + \eta_t^*.
$$

Estimation of the first and third row will be called the first step regression. The estimates for $\alpha_{1,1}$ and $\alpha_{3,1}$ would then be used for causality interpretation. Finding them to correspond to a nonzero ($\alpha_{1,1}$) matrix and to a zero ($\alpha_{3,1}$) matrix

respectively would yield the correct interpretation, namely that in the long run causality is running from z to y but not vice versa. Let

$$Y = \begin{bmatrix} \Delta y_1 & \Delta y_2 & \cdots & \Delta y_t \\ \Delta z_1 & \Delta z_2 & \cdots & \Delta z_t \end{bmatrix}, X' = \begin{bmatrix} (y_0 - \Lambda^* z_0)' \\ (y_1 - \Lambda^* z_1)' \\ \vdots \\ (y_{t-1} - \Lambda^* z_{t-1})' \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_{1,1} \\ \alpha_{3,1} \end{bmatrix}, \bar{\eta}_t = \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix},$$

and obtain the regression model $Y = \alpha X + \bar{\eta}_t$ with ordinary least squares (OLS) estimates $\hat{\alpha} = YX'(XX')^{-1}$. Giving the set of elements L which solve $det(I_{n_2} - \Phi L) = 0_{n_2 \times n_2}$ obtains

$$\Delta z_t = \begin{cases} \tilde{\eta}_{2,t} \equiv \eta_{2,t} \in L \\ \tilde{\eta}_{2,t} \equiv -\Psi(\Phi^{t-1} z_0 + \sum_{i=0}^{t-1} \Phi^i \eta_{2,i}^{t-i}) + \eta_{2,t} \in L, \end{cases}$$

$$\Delta y_t = \Lambda \tilde{\eta}_{2,t+1} + \Lambda_\Delta \iota_{t+1} + \eta_{1,t} - \eta_{1,t-1},$$

$$y_{t-1} - \Lambda^* z_{t-1} = \eta_{1,t-1} + \Lambda \iota_{t+1} + \eta_{2,t}^\phi$$

$$\eta^\phi = \Lambda \sum_{i=t+1-k}^{t+1} \Phi^{t+1-i} \eta_{2,i}, \ k = 1 \text{ (in general: } k \text{ finite)}.$$

The following definitions help to simplify the representation. Assume that the model is constant over time and that the innovations $\eta_{i,t}$, $i = 1,2$ as well as the forecast error, ι_t has time invariant first and second moments.[8] We then let for a covariance stationary stochastic variable $W = (w_1, w_2, \ldots, w_t)$

$$\text{plim}(WW'/T) = \Sigma_{w_s} = \frac{1}{T} \sum_{t=1}^{T} w_t w_{t+s}', \ |s| = 0,1,2,\ldots$$

8   Note that stationarity of the forecast error is part of the necessary conditions for writing (1) as a cointegrated system.

be the probability limit for the covariance estimator of the covariance between w_t and w_{t+s}.[9] Furthermore we may note that $\Sigma_{w_{s-j}} = \Sigma_{w_{s+j}}$, $|s|,|j|=1,2,\ldots$ Similar arguments hold for w_t and another covariance stationary stochastic variable τ_t:

$$\Sigma_{w,\tau_s} = \frac{1}{T}\sum_{t=1}^{T} w_t \tau'_{t+s}, \quad |s|=1,2,\ldots$$

The asymptotic OLS biases are given as

$$a_1 = \operatorname{plim}(\hat{\alpha}_{1,1} - \alpha_{1,1}) \text{ and } a_2 = \operatorname{plim}(\hat{\alpha}_{3,1} - \alpha_{3,1}),$$

the corresponding formulas being

$$a_i = \operatorname{plim}\left(\frac{\eta_{i,t} X'}{T}\right)\operatorname{plim}\left(\frac{XX'}{T}\right)^{-1}, \ i=1,3.$$

It then follows that

$$\operatorname{plim}(XX'/T) = \left[\Sigma_{\eta_1} + 2\Sigma_{\eta_1 \iota_2}\Lambda' + \Lambda\Sigma_{\iota}\Lambda' + 2\Sigma_{\eta_2^\phi \iota_1}\Lambda' + \Sigma_{\eta_2^\phi} + 2\Sigma\eta_1\eta_{2,1}^\phi\right] \quad (7)$$

and

$$a_1 = \begin{bmatrix} \Sigma_{\eta_{1,1}} + \Sigma_{\eta_1 \iota_2}\Lambda' + \Sigma_{\eta_1 \eta_{2,2}}\Lambda' + \Sigma_{\eta_1 \eta_{2,1}}\Phi'\Lambda' \\ +\Sigma_{\eta_1 \iota_1}\Lambda' + \Lambda\Sigma_{\iota}\Lambda' + \Lambda\Sigma_{\eta_2 \iota}\Lambda' + \Lambda\Sigma_{\eta_2 \iota}\Phi'\Lambda' \\ +\Sigma_{\eta_2^\phi \eta_1} + \Sigma_{\eta_2^\phi \iota_1}\Lambda' + \Sigma_{\eta_2^\phi \eta_{2,1}}\Lambda' + \Sigma_{\eta_2^\phi \eta_2}\Phi'\Lambda' \end{bmatrix}' \\ \times [\Sigma_{\eta_1} + 2\Sigma_{\eta_1 \iota_2}\Lambda' + \Lambda\Sigma_{\iota}\Lambda' + 2\Sigma_{\eta_2^\phi \iota_1}\Lambda' + \Sigma_{\eta_2^\phi} + 2\Sigma\eta_1\eta_{2,1}^\phi]^{-1}.$$

Both factors on the right hand side involve the term $\Lambda\Sigma_{\iota}\Lambda'$ which implies that a_1 cannot be assumed to turn out zero. In fact, since $\Lambda\Sigma_{\iota}\Lambda'$ is positive semi-definite under reasonable circumstances there is a tendency for the $\hat{\alpha}_{1,1}$ to be biased upward. Hence, it will in general not be informative about the true causal links.

Turning to a_3, we find

$$a_3 = [\Sigma_{\eta_1 \bar{\eta}_{2,1}} + \Lambda\Sigma_{\iota \bar{\eta}_{2,1}} + \Sigma_{\bar{\eta}_2 \eta_2^\phi}]' \\ \times [\Sigma_{\eta_1} + 2\Sigma_{\eta_1 \iota_2}\Lambda' + \Lambda\Sigma_{\iota}\Lambda' + 2\Sigma_{\eta_2^\phi \iota_1}\Lambda' + \Sigma_{\eta_2^\phi} + 2\Sigma\eta_1\eta_{2,1}^\phi]^{-1}.$$

---

9   For $s = 0$ the subscript will be omitted.

To interpret z to not depend on y the estimate of $\hat{\alpha}_{3,1}$ should be a matrix of zeros. This again will generally not be the case. The reason is the term

$$\Sigma_{\tilde{\eta}_2 \eta_2^{\phi}}$$

in the first factor on the right hand side which is not going to be zero even if the remaining matrix expressions which involve cross covariances might do. Moreover, it would not even help to find $\iota_t = 0$, $\forall t$ (perfect forecasts). In that situation the most reasonable effect would be an even larger bias since the 'denominator', (XX'/T), would be 'smaller' and hence the bias would not be reduced as much as it is due to $\Lambda \Sigma_\iota \Lambda'$ for non-zero forecast errors. On the other hand, if the forecast error had a huge variance in comparison to the innovations in η_i the bias would disappear. This effect gives rise to what has been called 'good forecast bias'. Similarly, for very large Σ_{η_1} implying that the economic model does not make much sense, the bias would also disappear. Therefore, the true causal links between y and z will be obtained only if the underlying economic model is poor as defined in the main text.

A.2 Step 2 – The Complementary Regression

The standard regression approach appeared to produce unreliable or even totally misleading results with respect to the coefficients of interest. Therefore, a complementary regression has been suggested in Section 4. Consider

$$\tilde{A}_0 Y_t \;=\; \tilde{A}_1 Y_{t-1} + \tilde{\eta}$$

$$\begin{bmatrix} I_{n_1} & -\Lambda \\ 0_{n_2 \times n_1} & I_{n_2} \end{bmatrix} \begin{bmatrix} y_t \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} 0_{n_1 \times n_1} & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & \Phi \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_t \end{bmatrix} + \begin{bmatrix} \eta_{1,t} + \Lambda \iota_{t+1} \\ \eta_{2,t+1} \end{bmatrix},$$

Pre-multiplying with the inverse of $\tilde{A}_0$ and subtracting Y_{t-1} from both sides gives

$$\Delta \begin{bmatrix} y_t \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} -I_{n_1 \times n_1} & \Lambda \Phi \\ 0_{n_2 \times n_1} & -\Psi \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_t \end{bmatrix} + \tilde{A}_0^{-1} \begin{bmatrix} \eta_{1,t} + \Lambda \iota_{t+1} \\ \eta_{2,t+1} \end{bmatrix}$$

which reproduces the triangular structure seen before and which ensures that the economic and Granger causality coincide also in this partial model. The decomposition of the matrix in front of the lagged right hand side variables follows the lines above and the coefficient of interest will be called $\alpha_{1,1}$ and $\alpha_{2,1}$ respectively. The new regressor, $\tilde{X}$, is now given by

$$\tilde{X}' = \begin{bmatrix} (y_0 - \Lambda\Phi z_1)' \\ (y_1 - \Lambda\Phi z_2)' \\ \vdots \\ (y_{t-1} - \Lambda\Phi z_t)' \end{bmatrix},$$

giving rise to

$$\text{plim}\left(\frac{\tilde{X}\tilde{X}'}{T}\right) = \begin{bmatrix} \Lambda\Sigma_{\tilde{\eta}_2}\Lambda' - 2\Lambda\Sigma_{\tilde{\eta}_2\eta_2}\Lambda' - 2\Lambda\Sigma_{\tilde{\eta}_2\eta_{1,1}} - 2\Lambda\Sigma_{\tilde{\eta}_2\iota}\Lambda' \\ +\Lambda\Sigma_{\eta_2}\Lambda' + 2\Sigma_{\eta_1\eta_{2,1}}\Lambda' + 2\Lambda\Sigma_{\eta_2\iota}\Lambda' \\ +\Sigma_{\eta_1} + 2\Sigma_{\eta_1\iota_1}\Lambda' \\ +\Lambda\Sigma_\iota\Lambda' \end{bmatrix}, \tag{8}$$

which can be simplified to yield

$$\text{plim}(\tilde{X}\tilde{X}'/T\big|_{\Phi = I_{n_2}}) = [\Sigma_{\eta_{1,0}} + 2\Sigma_{\eta_1\iota_1}\Lambda' + \Lambda\Sigma_\iota\Lambda'],$$

in case of $\Phi = I_{n_2}$ which corresponds to the cointegration approach pursued in the application. Furthermore, the biases $\tilde{a}_1$, $\tilde{a}_2$ for the estimates $\alpha_{1,1}$ and $\alpha_{2,1}$ are now

$$\tilde{a}_1 = \begin{bmatrix} -\Lambda\Sigma_{\tilde{\eta}_2\eta_{1,1}}\Lambda' - \Lambda\Sigma_{\tilde{\eta}_2\iota_1}\Lambda' - \Lambda\Sigma_{\tilde{\eta}_2\eta_{2,1}}\Lambda' \\ +\Lambda\Sigma_{\eta_2\eta_{1,1}}\Lambda' + \Lambda\Sigma_{\eta_2\iota_1}\Lambda' + \Lambda\Sigma_{\eta_2\eta_{2,1}}\Lambda' \\ +\Sigma_{\eta_{1,1}} + \Sigma_{\eta_1\iota_2}\Lambda' + \Sigma_{\eta_1\eta_{2,2}}\Lambda' \\ +\Lambda\Sigma_{\iota\eta_1} + \Lambda\Sigma_{\iota_1}\Lambda' + \Lambda\Sigma_{\iota\eta_{2,1}}\Lambda' \end{bmatrix} \times \text{plim}\left(\frac{\tilde{X}\tilde{X}'}{T}\right)^{-1},$$

$$\tilde{a}_2 = [\Lambda\Sigma_{\tilde{\eta}_2\eta_{2,1}} - \Lambda\Sigma_{\tilde{\eta}_{2,1}} + \Sigma_{\eta_1\tilde{\eta}_{2,2}} + \Lambda\Sigma_{\tilde{\eta}_2\iota_1}] \times \text{plim}\left(\frac{\tilde{X}\tilde{X}'}{T}\right)^{-1}.$$

Again, one might look at the special case for $\Phi = I_{n_2}$. We then have

$$\tilde{a}_1\big|_{\Phi=I_{n_2}} = \left[\begin{matrix} \Sigma_{\eta_{1,1}} + \Sigma_{\eta_1 \iota_2}\Lambda' + \Sigma_{\eta_1 \eta_{2,2}}\Lambda' \\ + \Lambda\Sigma_{\iota \eta_1} + \Lambda\Sigma_{\iota_1}\Lambda' + \Lambda\Sigma_{\eta_{2,1}}\Lambda' \end{matrix}\right] \times \text{plim}\left(\tilde{X}\tilde{X}'/T\big|_{\Phi=I_{n_2}}\right)^{-1},$$

$$\tilde{a}_2\big|_{\Phi=I_{n_2}} = \left[\Sigma_{\eta_1 \tilde{\eta}_{2,2}} + \Lambda\Sigma_{\tilde{\eta}_{2,1}}\right] \times \text{plim}\left(\tilde{X}\tilde{X}'/T\big|_{\Phi=I_{n_2}}\right)^{-1}.$$

It is thus straightforward to see that for smaller forecast errors ($\iota_t \approx 0_{n_2\times 1}$), $\tilde{a}_2$ will be closer to zero than otherwise. The same holds for small variations in $\eta_{1,t}$. Under the same conditions $\hat{\alpha}_{1,1}$ will approach $\alpha_{1,1}$.[10] Thus, the better the economic model, the higher is the chance that the true causal links will be revealed. However, it is not known a priori, if one finds herself or himself in the standard regression or in the complementary regression. That's why the following procedure can be suggested.

1. Run a regression as in the standard case.
2. Lead the set of regressors which do not turn out weakly exogenous to the most likely period for which expectations of these variables may count for the weakly exogenous variables.
3. Run the complementary regression.
4. If the same set of variables turns out weakly exogenous as before they can be considered the driving variables.
5. If the set of variables that has previously been found to be endogenous turns out weakly exogenous, lead the weakly exogenous variables of the first step appropriately and run another regression.
6. If the set of variables turning out weakly exogenous is the same as in step 2, then they can be considered the driving variables

Otherwise no set of variables can be labelled causal. The number 5 of the procedure above could be regarded a third step regression, but in fact it is merely a confirmation of the correct choice and could also be omitted. In Table 2 the results for this third regression are reported in Section 6.

---

10 Independent of Φ the 'numerators' now only contain cross terms which would vanish if zero correlation between $\eta_{1,t}$ and $\eta_{2,r}$ for all r and t is assumed and if the corellation between $\eta_{1,t}$, $\eta_{1,r}$, ι_t, and ι_r is zero for all $r \neq t$.

A.3 Generalisation for $s \geq 1$

So far, s has been restricted to equal 1. It is easy to see however, that the results can be generalised to any value of s. Consider

$$\tilde{y}_t = \begin{bmatrix} y_t \\ \vdots \\ y_{t-s} \end{bmatrix}, \ \tilde{z}_t = \begin{bmatrix} z_{t+s-1} \\ \vdots \\ z_{t-1} \end{bmatrix}, \ \Lambda^+ = I_s \otimes \Lambda, \ \Phi^+ = I_s \otimes \Phi.$$

Then replace Λ by Λ^+, Φ by Φ^+, y_t by $\tilde{y}_t$ and z_t by $\tilde{z}_t$ in (1) and the analysis goes through.

B.   Data Sources

Table 3. Data Descriptions and Data Sources

Model (Table 1)	Item / description	Code	Source
1	CPI infl.: 1200-fold log of 1st difference of Consumer price index, all items less food and energy Base Period: 1982–84=100, seasonally adjusted with X12Arima	CUUR0000SA0L1E	USA, bureau of labor statistics (BLS)
	Bond y.: Rate of interest in money and capital markets, Federal Government securities, Constant maturity Ten-years		Federal Reserve System (FED)
2	CPI infl.: Switzerland, 1200-fold log of 1st difference of Consumer price index, all items Base Period: May 1993=100, seasonally adjusted	TS11515102	Switzerland, Federal Bureau of Statistics
	Bond y.: Switzerland, Rate of interest in Federal Government securities, Constant maturity Ten-years		Swiss National Bank (SNB), Monthly Bulletin (MB) 08/2003, Table E3
3	LIBOR: Germany, Money Market Rate, 3 months	SU0107	Bundesbank, MB 08/2003
	LIBOR: Switzerland, Money Market Rate, 3 months		SNB, MB 08/2003
4	Bond y.: USA see Model 1		

Model (Table 1)	Item / description	Code	Source
	Bond y.: Japan, Government Bond Yield	M.15861…ZF…	International Monetary Fund (IMF)
	LIBOR: USA Eurodollar deposits, Primary market, three-month maturity		FED
	LIBOR: Japan, 3-MONTH LIBOR: Offer London	M.15860EA.ZF…	IMF

References

ENGLE, RICHARD F., and CLIVE W. J. GRANGER (1987), "Co-Integration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55 (2), pp. 251–276.

ERICSSON, NEILL RICHARDS (1992), "Cointegration, Exogeneity, and Policy Analysis: An Overview", *Journal of Policy Modeling*, 14, pp. 251–280.

ERICSSON, NEILL RICHARDS, DAVID F. HENDRY, and GRAYHAM E. MIZON (1998), "Exogeneity, Cointegration, and Economic Policy Analysis", *Journal of Business and Economic Statistics*, 16 (4), pp. 370–387.

FISHER, IRVING (1930), *The Theory of Interest*, New York.

GRANGER, CLIVE W. J., and J. LIN (1995), "Causality in the Long Run", *Econometric Theory*, 11 (1), pp. 530–536.

HENDRY, DAVID F., and GRAYHAM E. MIZON (1998), "Exogeneity, Causality, and Cobreaking in Economic Policy Analysis of a Small Econometric Model of the Money in the UK", *Empirical Economics*, 23 (3), pp. 267–294.

HOSOYA, YUZO (1991), "The Decomposition and Measurement of the Interdependence between Second-Order Stationary Processes", *Probalility Theory and Related Fields*, 88, pp. 429–444.

HUBRICH, KIRSTIN (2001), *Cointegration Analysis in a German Monetary System*, Heidelberg.

JOHANSEN, SØREN (1988), "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics*, 12, pp. 231–254.

JOHANSEN, SØREN (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 1st ed., Oxford.

JOHANSEN, SØREN, and KATARINA JUSELIUS (1990), "Maximum Likelihood Estimation and Inference on Cointegration – With Applications to the Demand for Money", *Oxford Bulletin of Economics and Statistics*, 52, pp. 169–210.

JUSELIUS, KATARINA (1996), "An Empirical Analysis of the Changing Role of the German Bundesbank after 1983", *Oxford Bulletin of Economics and Statistics*, 58, pp. 791–819.

JUSELIUS, KATARINA, and RONALD MacDONALD (2004), "International Parity Relationships Between the USA and Japan", *Japan and the World Economy*, 16, pp. 17–34.

LÜTKEPOHL, HELMUT, and JÜRGEN WOLTERS (2003), "Transmission of German Monetary Policy in the Pre-Euro Period", *Macroeconomic Dynamics*, 7 (5), pp. 711–733.

PESARAN, M. H., YONGCHEOL SHIN, and RICHARD J. SMITH (2000), "Structural Analysis of Vector Error Correction Models with Exogenous I(1) Variables", *Journal of Econometrics*, 97, pp. 293–343.

SUMMARY

This paper considers an expectations augmented version of the ENGLE and GRANGER (1987) error correction model and shows that standard inference about the adjustment coefficients can be severely biased. This bias has implications for long–run causality and impulse–response analysis in particular. However, a sometimes simple remedy exists which only requires some additional regressions. The results are illustrated with popular macroeconomic relationships like the Fisher relation and uncovered interest parity hypothesis using U.S., German and Swiss data.