# The long swings puzzle: what the data tell when allowed to speak freely

Katarina Juselius
Department of Economics, University of Copenhagen
First version: October 13, 2007
this version: October 27, 2008

**Abstract**

The persistent movements away from long-run benchmark values in real exchange rates often observed in periods of currency float have been subject to much empirical research without resolving the underlying theoretical puzzle. This chapter demonstrates how the Cointegrated VAR approach of grouping together components of similar persistence can be used to uncover structures in the data that ultimately may help to explain theoretically the forces underlying such puzzling movements. The charaterization of the data into components which are empirically I(0), I(1) and I(2) is shown to be a powerful organizing principle, allowing us to structure the data into long-run, medium-run, and short-run behavior. Its main advantage is the ability to associate persistent movements away from fundamental benchmark values in one variable/relation with similar persistent movements somewhere else in the economy.

JEL: C32, C50, F41

Keywords: Cointegrated VAR, I(2), Deterministic componenets, Persistent movements.

# Introduction

International macroeconomics is known for a number of empirical puzzles, the most notable among them being the 'PPP puzzle', which is closely related to the 'long swings puzzle' and the 'exchange disconnect puzzle' (Rogoff, 1996). These puzzles are all related to the pronounced persistence away from equilibrium states that have been observed in many real exchange rates during periods of currency float. Among these, the Dmk-$ rate in the post Bretton Woods period is one of the more extreme cases.

One important purpose of this chapter is to demonstrate how the Cointegrated VAR (CVAR) approach (Juselius, 2006) can be used to uncover structures in the data that ultimately may help to explain theoretically the forces underlying such persistent movements in the data. The CVAR approach starts from a general unrestricted VAR model that gives a good characterization of the the raw data. It then tests down until a parsimonious representation of the data with as much economic content as possible has been achieved. When properly applied, the CVAR is able to extract valuable information about the dynamics of the pulling and pushing forces in the data without distorting this information. This entails the identification of stationary relationships between nonstationary variables, interpretable as long-run equilibrium states, and the dynamic adjustment of the system to deviations from these states. It also entails the identification of the transitory and permanent shocks that have affected the variables and the short-run and long-run impact of these shocks.

For the results to be reliable, the statistical properties of the model, have, however, to be taken seriously. This implies adequately controlling for reforms, interventions, regime shifts, etc., that often are part of the data generating mechanism. The reunification of East and West Germany is an example of such an important event. The approach also entails the untying of any transformation of the variables, such as the *real exchange* transformation, imposed from the outset on the data. Such transformations, common in empirical economics, often seriously distort signals in the data that, otherwise, might help to uncover precisely those empirical regularities which give a clue to the underlying reasons for the puzzling behavior.

The weight of the empirical analysis is on characterizing data within the broad framework of a theory model. To facilitate the interpretation of the empirical results, the chapter argues that it is essential first to translate the underlying assumptions of the theoretical model into hypotheses on the pulling and pushing forces of the VAR model (Juselius and Johansen, 2006, Juselius, 2006, Juselius and Franchi, 2007). A careful formulation of such a scenario is indispensable for being able to structure and interpret the emprical results so that empirical regularities either supporting or rejecting the theoretical assumptions become visible. In particular, the latter are valuable as they should ultimately lead to empirically more relevant theory models. Thus, to some extent, the CVAR approach switches the role of theory and statistical analysis in the sense of rejecting the privileging of *a priori* economic theory over empirical evidence. In the language of the CVAR approach, empirical evidence is the pushing force and economic theory is adjusting (Hoover et al., 2007)

The approach will be illustrated with an empirical analysis of the long swings in real exchange rates based on German and US prices and the Dmk/$ rate over the period 1975:09-1998:12. Using the above decomposition into pulling and pushing forces, the empirical analysis identifies a number of 'structured' (rather than stylized) facts describing important empirical regularities underlying the long swings puzzle. These provide clues suggesting where to dig deeper (see Hoover, 2006) to gain an empirically more relevant understanding of the puzzling behavior in the goods and foreign exchange markets.

To structure the data as efficiently as possible, this chapter argues that the order of integration, rather than a structural parameter, should be considered an empirical approximation, measuring the degree of persistent behavior in a variable or a relation. Organizing the data into directions where they are *empirically* $I(0)$, $I(1)$ or $I(2)$ is not the same as claiming they are *structurally* $I(0)$, $I(1)$, or $I(2)$. In the first case some implications of the statistical theory of integrated processes are likely to work very well, such as inference on structures, others less well, such as inference on the long-run values to which the process converges towards when all the errors have been switched off (Johansen, 2005). The focus of this chapter is on structure rather than long-run values.

The statistical analysis suggested that the two prices (and possibly even the nominal exchange rate) were empirically $I(2)$. Thus, another important aim of this chapter is to discuss the $I(2)$ model, how it relates to the $I(1)$ model, and what can be gained by interpreting the empirical reality within the rich structure of the $I(2)$ model. Because the $I(2)$ model is also more complex, the analysis is first done in the $I(1)$ model, emphazing those signals in the results suggesting data are $I(2)$. Though most of the $I(1)$ results can be found in the $I(2)$ model, the chapter demonstrates that the $I(2)$ results are more precise and that the $I(2)$ structure allows for a far richer interpretation.

The exposition of the chapter is as follows: Section 1 defines the $I(1)$ and $I(2)$ models as parameter restrictions on the unrestricted VAR. Section 2 introduces the persistent features of the *real exchange* data for the German-US case and discusses how they can be formulated in the pulling and pushing forces of a CVAR model. Section 3 discusses under which conditions $I(2)$ data can be modelled with the $I(1)$ model, why it works, and how the interpretation of the results has to be modified. Section 4 presents the empirical $I(1)$ analysis of prices and nominal exchange rates inclusive of specification testing and estimation of the long-run structure. Section 5 gives a brief account of the $I(2)$ model and discusses at some length the specification of the deterministic componenets. Section 6 discusses an estimation procedure based on maximum likelihood and shows how the $I(2)$ structure can be linked to the $I(1)$ model. Section 7 provides a theoretical scenario for the *real exchange* data. Section 8 presents the empirical results of the pulling and pushing forces structured by the $I(2)$ model, summarizes the puzzling facts detected, and discusses what has been gained by this analysis compared to the $I(1)$ analysis. Section 9 concludes with a discussion of what the data were able to tell when allowed to speak freely.

# 1 The VAR model

The baseline VAR(2) model in its unrestricted form is given by:

$$\mathbf{x}_t = \mathbf{\Pi}_1 \mathbf{x}_{t-1} + \mathbf{\Pi}_2 \mathbf{x}_{t-2} + \mathbf{\Phi} \mathbf{D}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

with

$$\boldsymbol{\varepsilon}_t \sim N_p(\mathbf{0}, \mathbf{\Omega}\ ),\ t = 1, ..., T$$

where $\mathbf{x}_t' = [x_{1,t}, x_{2,t}, ...x_{p,t}]$ is a vector of $p$ stochastic variables and $\mathbf{D}_t$ is a vector of deterministic variables, such as constant, trend and various dummy variables. As the subsequent empirical VAR model has lag two, all results are given for the VAR(2) model. A generalization to higher lags should be straightforward.

In terms of likelihood, an equivalent formulation of (1) is the vector equilibrium correction form:

$$\Delta \mathbf{x}_t = \mathbf{\Gamma}_1 \Delta \mathbf{x}_{t-1} + \mathbf{\Pi} \mathbf{x}_{t-1} + \mathbf{\Phi} \mathbf{D}_t + \boldsymbol{\varepsilon}_t, \tag{2}$$

where $\mathbf{\Gamma}_1 = -\mathbf{\Pi}_2$ and $\mathbf{\Pi} = -(\mathbf{I} - \mathbf{\Pi}_1 - \mathbf{\Pi}_2)$.

Alternatively, (1) can be formulated in acceleration rates, changes and levels:

$$\Delta^2 \mathbf{x}_t = \mathbf{\Gamma} \Delta \mathbf{x}_{t-1} + \mathbf{\Pi} \mathbf{x}_{t-1} + \mathbf{\Phi} \mathbf{D}_t + \boldsymbol{\varepsilon}_t, \tag{3}$$

where $\mathbf{\Gamma} = -(\mathbf{I} - \mathbf{\Gamma}_1)$ and $\mathbf{\Pi}$ as above. As long as all parameters are unrestricted, the VAR model is no more than a convenient summary of the covariances of the data. As a result, most VAR models are heavily overparametrized and insignificant parameters need to be set to zero. The idea of general-to-specific modelling is to reduce the number of parameters by significance testing, with the final aim of finding a parsimonious parameterization with interpretable economic contents. Provided that the simplification search is statistically valid, the final restricted model will reflect the full information of the data. Thus, given the broad framework of a theory model, a correct CVAR analysis allows the data to speak freely about the underlying mechanisms that have generated the data.

All three models are equivalent from a likelihood point of view, but (1) would generally be chosen when $\mathbf{x}_t$ is $I(0)$, (2) when $\mathbf{x}_t$ is $I(1)$, and (3) when $\mathbf{x}_t$ is $I(2)$. The hypothesis that $\mathbf{x}_t$ is $I(1)$ is formulated as a reduced rank restriction on the matrix $\mathbf{\Pi}$:

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' \text{ , where } \boldsymbol{\alpha}, \boldsymbol{\beta} \text{ are } p \times r, \tag{4}$$

and that $\mathbf{x}_t$ is $I(2)$ as an additional reduced rank restriction on the transformed matrix $\mathbf{\Gamma}$:

$$\boldsymbol{\alpha}'_{\perp}\mathbf{\Gamma}\boldsymbol{\beta}_{\perp} = \boldsymbol{\xi}\boldsymbol{\eta}', \text{ where } \boldsymbol{\xi}, \boldsymbol{\eta} \text{ are } (p-r) \times s_1, \tag{5}$$

where $\boldsymbol{\beta}_{\perp}, \boldsymbol{\alpha}_{\perp}$ are the orthogonal complements of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. The first reduced rank condition is formulated on the variables in levels, the second on the variables in differences. The intuition of the former is that the variables contain stochastic trends (unit roots) that can be cancelled by linear combinations. The intuition of the latter is that the differenced process also contains unit roots when data are $I(2)$. However, in this case the linear combinations that cancel these roots are more complicated. Thus, when $\mathbf{x}_t \sim I(2)$ and, hence $\Delta\mathbf{x}_t \sim I(1)$, it is not sufficient to impose the reduced rank restriction on the matrix $\mathbf{\Pi}$ to get rid of all (near) unit roots in the model. This is because $\Delta\mathbf{x}_t$ is also a unit root process and lowering the value of $r$ does not remove the unit roots belonging to $\mathbf{\Gamma} = -(\mathbf{I} - \mathbf{\Gamma}_1)$. Therefore, even though the rank of $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ has been correctly determined, there will remain additional unit roots in the VAR model when the data are $I(2)$. As will be demonstrated below, this provides a good diagnostic tool for detecting $I(2)$ problems in the VAR analysis.

Inverting the VAR model gives us the Moving Average (MA) form. Under the reduced rank of (4) and the full rank of (5), the MA form is given by:

$$\mathbf{x}_t = \mathbf{C}\sum_{i=1}^{t}(\boldsymbol{\varepsilon}_i + \mathbf{\Phi}\mathbf{D}_i) + \mathbf{C}^*(L)(\boldsymbol{\varepsilon}_t + \mathbf{\Phi}\mathbf{D}_t) + \mathbf{A} \tag{6}$$

where $\mathbf{C}^*(L)$ is a lag polynomial describing the impulse response functions of the empirical shocks to the system, and $\mathbf{A}$ is a function of the initial values $\mathbf{x}_0, \mathbf{x}_{-1}, \mathbf{x}_{-2}$, and $\mathbf{C}$ is of reduced rank $p-r$:

$$\mathbf{C} = \boldsymbol{\beta}_{\perp}(\boldsymbol{\alpha}'_{\perp}\Gamma\boldsymbol{\beta}_{\perp})^{-1}\boldsymbol{\alpha}'_{\perp} = \tilde{\boldsymbol{\beta}}_{\perp}\boldsymbol{\alpha}'_{\perp} \tag{7}$$

with $\tilde{\boldsymbol{\beta}}_{\perp} = \boldsymbol{\beta}_{\perp}(\boldsymbol{\alpha}'_{\perp}\boldsymbol{\beta}_{\perp})^{-1}$.

Inverting the VAR under the reduced rank of both (4) and (5) will be discussed in Section 5.

# 2 The persistent swings in real exchange data

Parity conditions are central to international finance and, more specifically, to many open economy macro-models, such as the popular Dornbusch (1976) sticky price overshooting model with rational expectations (RE). However, the persistent movements away from long-run benchmark values that have characterized the Dmk/$ currency float of the post Bretton Woods period are hard to reconcile with the Dornbush model and its many modifications (see Frydman, Goldberg, Johansen and Juselius (2008) and references therein). They show that it is, in particular, the assumption of RE in these models that is inconsistent with the long swings behavior characterizing many real exchange rates. This prompts for a reformulation of the model and of how agents are assumed to make forcasts. The idea here is to formulate the basic assumptions of the PPP hypotheses under a currency float into testable hypotheses on the pushing and pulling forces of the Cointegrated VAR model, a so called scenario. By comparing assumed with actual behavior it should be possible to pinn down exactly where the puzzling behavior is. Since the VAR model is just a reformulation of the covariance information in the data, the end results should be a set of empirical features which a theory model should be able to replicate in order to claim empirical relevance.

## 2.1 The long swings puzzle

Purchasing power parity ($PPP$) is defined as:

$$p_1 = p_2 + s_{12}, \tag{8}$$

where $p_1$ is the log of the domestic price level (here German), $p_2$ is the log of the foreign price level (here US), and $s_{12}$ denotes the log of the spot exchange rate (here Dmk/$). The real exchange rate, $ppp_t$, is the departure at time $t$ from (8):

$$ppp_t = p_{1,t} - p_{2,t} - s_{12,t}. \tag{9}$$

An ocular inspection gives a first impression of the development over time of prices and the nominal exchange rate and illustrates what the puzzle is about. Figure 1, upper panel, shows that US prices have grown more than German prices resulting in a downward sloping stochastic trend in relative prices. According to purchasing power parity, the nominal exchange rate should reflect this downward sloping trend. The picture shows that this is also approximately the case over the very long run. However, what is striking are the long swings around that downward sloping trend.

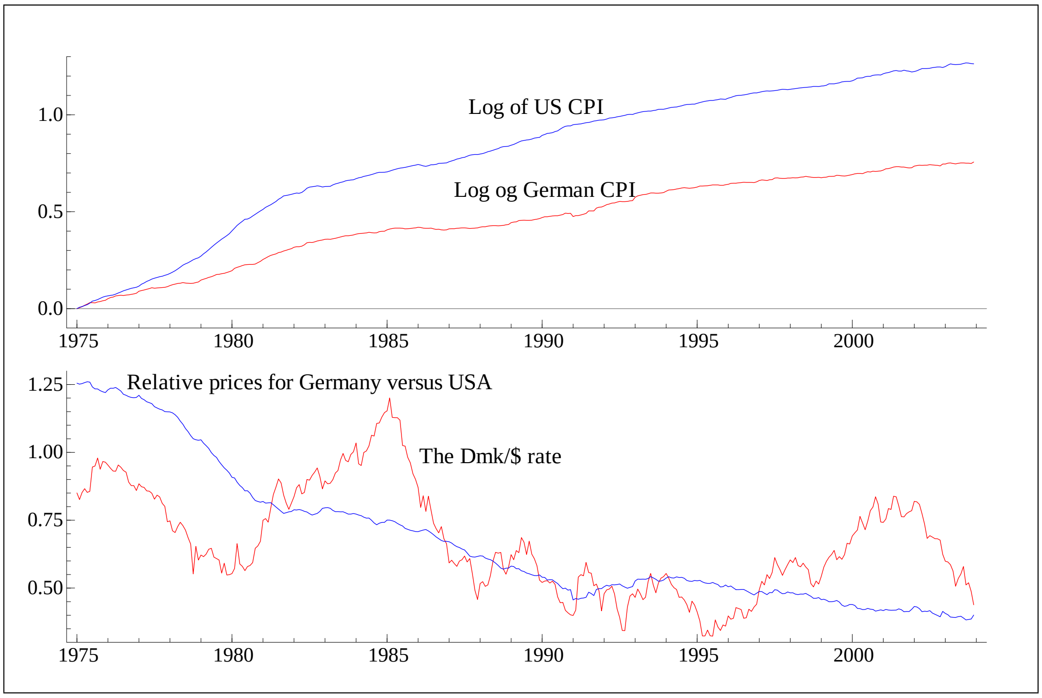


Figure 1: The time graphs of German and US prices (upper panel) and their relative prices and nominal exchange rate (lower panel.

How can we use econometrics to learn about the mechanisms underlying these swings? The subsequent VAR analyis will demonstrate that the joint modelling of prices and exchange rates allows us to formulate much richer hypotheses about the empirical mechanisms behind the puzzle.

## 2.2 Pulling and pushing forces in the cointegrated VAR model

To provide the intuition for the VAR approach and to show how the results can be interpreted in terms of pulling and pushing forces, a hypothetical VAR analysis of the German-US PPP data will be used as an illustration. For simplicity, the discussion will be restricted to a bivariate $I(1)$ model for relative prices and the nominal exchange rate. Because the period of interest defines a currency float, a prior hypothesis is that the nominal exchange rate has been adjusting and prices pushing. Provided that the stochastic trend in nominal exchange rates reflects the stochastic trend in relative prices, it is easy to show that $ppp = p_1 - p_2 - s_{12} \sim I(0)$. Thus, the stationarity of PPP and its adjustment dynamics can be formulated as a composite hypothesis:

$\{(p_1 - p_2) = pp \sim I(1),\ s_{12} \sim I(1),\ ppp \sim I(0),\ s_{12}$ is adjusting, and $p_1, p_2$ are pushing$\}$.

The pulling forces are described by the vector equilibrium correction model:

$$\begin{bmatrix} \Delta pp_t \\ \Delta s_{12,t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} (pp_t - s_{12,t} - \beta_0) + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix},$$

where $(pp_t - s_{12,t} - \beta_0) = \boldsymbol{\beta}'\mathbf{x}_t$ is the cointegration relation with $E(ppp_t) = \beta_0$. Thus, an equilibrium position, defined as $pp_t - s_{12,t} = \beta_0$, can be given an interpretation as a resting point towards which the process is drawn after it has been pushed away. In this sense, an equilibrium position exists at all time points, $t$, contrary to the long-run value of the process, which is the value of the process in the limit as $t \to \infty$ and all shocks have been switched off.

The pushing forces are described by the corresponding common trends model:

$$\begin{bmatrix} pp_t \\ s_{12,t} \end{bmatrix} = \begin{bmatrix} c \\ c \end{bmatrix} \boldsymbol{\alpha}'_\perp \sum_{i=1}^{t} \boldsymbol{\varepsilon}_i + \mathbf{C}^*(L) \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix},$$

with $\boldsymbol{\alpha}'_\perp = \frac{1}{\alpha_1 - \alpha_2}[-\alpha_2, \alpha_1]$ and with $\boldsymbol{\alpha}'_\perp \sum_{i=1}^{t} \boldsymbol{\varepsilon}_i$ describing the common stochastic trend. Assume now that $\boldsymbol{\alpha}' = [0, \alpha_2]$, i.e. only the nominal exchange rate is equilibrium correcting when $ppp_t - \beta_0 \neq 0$. In this case $\boldsymbol{\alpha}'_\perp = [1, 0]$ implies that the common stochastic trend originates from relative price shocks. This would conform to the theoretical prior for a period of floating exchange rates.

The question is now whether the empirical reality given by the observed variables in Figure 1, lower panel, can be adequately represented by the above assumed pulling and pushing forces. Stationarity of $ppp_t$ would imply that the nominal exchange rate would follow relative prices one-for-one apart from stationary noise. Figure 2 shows a crossplot ofthe $pp_t$ and $s_{12,t}$ variables. If the assumption that $ppp_t \sim I(0)$ were correct, then the crossplots should be randomly scattered around the $45^0$ line defining the equilibrium position $pp_t = s_{12,t}$. Obviously, the crossplots measuring the deviation from $ppp$, i.e. $\boldsymbol{\beta}'\mathbf{x}_t = pp_t - s_{12,t} - \beta_0$, are systematically scattered either above or below the $45^0$ line. Thus the reality behind the observed real exchange rate looks very different from the assumed stationarity, illustrating the puzzle. The nonstationarity of real exchange rates has been demonstrated in a number of studies (see Froot and Rogoff, 1995, and MacDonald, 1995, for surveys; Cheung and Lai, 1993, Juselius, 1995, Johansen and Juselius, 1992, Juselius and MacDonald, 2004, 2007).

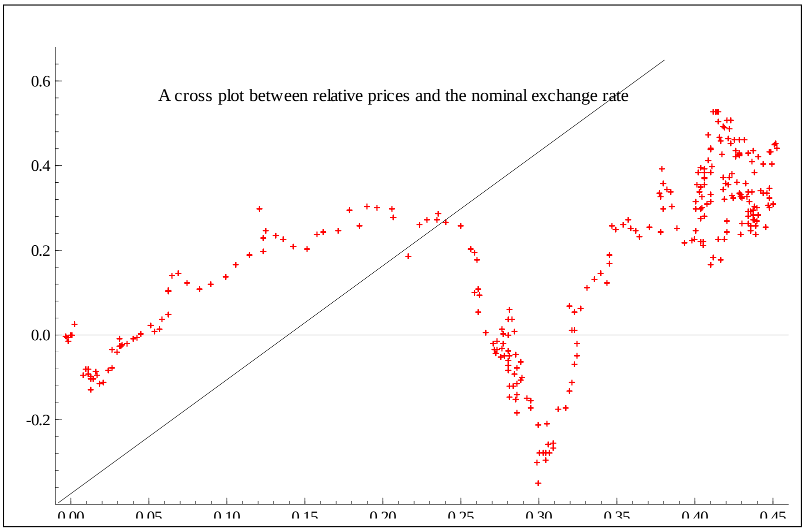


Figure 2: A cross plot of US-German relative prices and the \$/Dmk rate for the period 1975:4-1998:12.

## 2.3 Approximating persistent behavior with $I(1)$ or $I(2)$

The above ocular analysis showed that the long swings puzzle in PPP is essentially a question of why nominal exchange rates have so persistently moved away from relative prices. The previous subsection suggested that the cointegrated VAR model should be used to structure such data by the pulling and pushing forces. Section 2 defined the $I(1)$ and $I(2)$ models as reduced rank parameter restrictions on the $I(0)$ model, providing us with an empirically strong procedure for addressing behavioural macroeconomic problems. This is because the reduced rank parametrization of the CVAR allows us to group together components of similar persistence over the sample period. The charaterization of the data into *empirically* $I(0)$, $I(1)$ and $I(2)$ components is a powerful organizing principle, allowing us to structure the data in the long-run, medium-run, and short-run behavior. An additional advantage is that inference is likely to become more robust than otherwise. For example, treating a near unit root as stationary tends to invalidate certain inferences based on the $\chi^2$, $F$ and $t$ distributions unless we have a very long sample[1].

This is a fairly pragmatic way of classifying data allowing a variable to be treated as $I(1)$ in one sample and $I(0)$ or even $I(2)$ in another. The

[1] Johansen (2006) demonstrated that valid inference on steady-state values requires more than 5000 observations if the model contains a near unit root of 0.998.

idea is that, in a general equilibrium world, a persistent departure from a steady-state value of a variable or a relation should generate a similar persistent movement somewhere else in the economy. For example, if the Fisher parity holds as a stationary relation (stationary real interest rates) and we find that inflationary shocks have been very persistent, then we should expect interest rate shocks to have a similar persistence. Thus empirical persistence is a powerful property that can be used to investigate whether our prior hypothesis (the Fisher parity) is empirically relevant, and if not, which other variables have been co-moving in a similar manner, giving rise to new hypotheses.

From the outset, many economists would consider the idea that economic variables are $I(2)$ highly problematic. The argument is often that all inference on long-run values (the steady-state value a variable converges to when the errors are switched off) would lead to meaningless results. This is a valid argument provided one can argue that the order of integration is a structural parameter, which often seems doubtful. Nonetheless, there are cases when a structural interpretation is warranted. For example, Frydman *et al.* (2008) show that speculative behavior based on IKE is consistent with near I(2) behavior; arbitrage theory suggests that a nominal market interest rate should be a martingale difference process, i.e. approximately a unit root process. Of course, in such cases a structural (near) unit root should be invariant to the choice of sample period.

# 3 Modelling $I(2)$ data with the $I(1)$ model: does it work?

It often happens that $I(2)$ data are analyzed as if they were $I(1)$ because the $I(2)$ possiblity was never checked, or one might have realized that the data exhibit $I(2)$ features but decided to ignore these signals in the data. For this reason, it is of some interest to ask whether the findings from such $I(1)$ analyses are totally useless, misleading, or can be trusted to some extent.

Before answering these questions, it is useful to examine the so called **R**-model in which short-run effects have been concentrated out. We consider first the simple VAR(2) model:

$$\begin{gathered}\Delta\mathbf{x}_t = \mathbf{\Gamma}_1\Delta\mathbf{x}_{t-1} + \boldsymbol{\alpha\beta}'\mathbf{x}_{t-1} + \boldsymbol{\mu}_0 + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\varepsilon}_t \sim N_p(\mathbf{0}, \mathbf{\Omega}\ ),\ t = 1, ..., T\end{gathered} \tag{10}$$

and the corresponding **R**-model:

$$\mathbf{R}_{0t} = \boldsymbol{\alpha\beta}'\mathbf{R}_{1t} + \boldsymbol{\varepsilon}_t. \tag{11}$$

where $\mathbf{R}_{0t}$ and $\mathbf{R}_{1t}$ are found by concentrating out the lagged short-run effects, $\Delta\mathbf{x}_{t-1}$:

$$\Delta\mathbf{x}_t = \hat{\mathbf{B}}_1\Delta\mathbf{x}_{t-1} + \hat{\boldsymbol{\mu}}_0 + \mathbf{R}_{0t} \tag{12}$$

and

$$\mathbf{x}_{t-1} = \hat{\mathbf{B}}_2\Delta\mathbf{x}_{t-1} + \hat{\boldsymbol{\mu}}_0 + \mathbf{R}_{1t}. \tag{13}$$

When $\mathbf{x}_t \sim I(2)$, both $\Delta\mathbf{x}_t$ and $\Delta\mathbf{x}_{t-1}$ contain a common $I(1)$ trend which, therefore, cancels in the regression of one on the other as in (12). Thus, $\mathbf{R}_{0t} \sim I(0)$ even if $\Delta\mathbf{x}_t \sim I(1)$. On the other hand, an $I(2)$ trend cannot be cancelled by regressing on an $I(1)$ trend and regressing $\mathbf{x}_{t-1}$ on $\Delta\mathbf{x}_{t-1}$ as in (13) does not cancel the $I(2)$ trend, so $\mathbf{R}_{1t} \sim I(2)$. Because $\mathbf{R}_{0t} \sim I(0)$ and $\boldsymbol{\varepsilon}_t \sim I(0)$, equation (11) can only hold if $\boldsymbol{\beta} = \mathbf{0}$ or, alternatively, if $\boldsymbol{\beta}'\mathbf{R}_{1t} \sim I(0)$. Thus, unless the rank is zero, the linear combination $\boldsymbol{\beta}'\mathbf{R}_{1t}$ transforms the process from $I(2)$ to $I(0)$.

The connection between $\boldsymbol{\beta}'\mathbf{x}_{t-1}$ and $\boldsymbol{\beta}'\mathbf{R}_{1t}$ can be seen by inserting (13) into (11):

$$\begin{aligned}
\underbrace{\mathbf{R}_{0t}}_{I(0)} &= \boldsymbol{\alpha\beta}'(\underbrace{\mathbf{x}_{t-1}}_{I(2)} - \mathbf{B}_2\underbrace{\Delta\mathbf{x}_{t-1}}_{I(1)} - \hat{\boldsymbol{\mu}}_0) + \boldsymbol{\varepsilon}_t \\
&= \boldsymbol{\alpha}(\underbrace{\boldsymbol{\beta}'\mathbf{x}_{t-1}}_{I(1)} - \boldsymbol{\beta}'\underbrace{\mathbf{B}_2\Delta\mathbf{x}_{t-1}}_{I(1)} - \hat{\boldsymbol{\mu}}_0) + \boldsymbol{\varepsilon}_t \\
&= \boldsymbol{\alpha}(\underbrace{\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\omega}'\Delta\mathbf{x}_{t-1}}_{I(0)} - \hat{\boldsymbol{\mu}}_0) + \boldsymbol{\varepsilon}_t
\end{aligned} \tag{14}$$

where $\boldsymbol{\omega} = \boldsymbol{\beta}'\mathbf{B}_2$. It is now easy to see that the stationary relations $\boldsymbol{\beta}'\mathbf{R}_{1t}$ consist of two components $\boldsymbol{\beta}'\mathbf{x}_{t-1}$ and $\boldsymbol{\omega}'\Delta\mathbf{x}_{t-1}$. There are two possibilities:

1. $\boldsymbol{\beta}_i'\mathbf{x}_{t-1} \sim I(0)$ and $\boldsymbol{\omega}_i = \mathbf{0}$, where $\boldsymbol{\beta}_i$ and $\boldsymbol{\omega}_i$ denote the ith column of $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$, or

2. $\boldsymbol{\beta}_i'\mathbf{x}_{t-1} \sim I(1)$ cointegrates with $\boldsymbol{\omega}_i'\Delta\mathbf{x}_{t-1} \sim I(1)$ to produce the stationary relation $\boldsymbol{\beta}'\mathbf{R}_{1t} \sim I(0)$.

In the first case, we talk about directly stationary relations, in the second case about polynomially cointegrated relations. Here we shall consider $\boldsymbol{\beta}'\mathbf{x}_t \sim I(1)$ without distinguishing between the two cases, albeit recognizing that some of the cointegration relations $\boldsymbol{\beta}'\mathbf{x}_t$ may be stationary by themselves.

We have demonstrated above that $\mathbf{R}_{0t} \sim I(0)$ and $\boldsymbol{\beta}'\mathbf{R}_{1t} \sim I(0)$ in (11), which is the model on which all $I(1)$ estimation and test procedures are derived. This means that the $I(1)$ procedures can be used even though data are $I(2)$, albeit with the following reservations:

1. the $I(1)$ rank test cannot say anything about the reduced rank of the $\mathbf{\Gamma}$ matrix, i.e. about the number of $I(2)$ trends. The determination of the reduced rank of the $\mathbf{\Pi}$ matrix, though asymptotically unbiased, might have poor small sample properties (Nielsen and Rahbek, 2004)

2. the $\boldsymbol{\beta}$ coefficients relating $I(2)$ variables are $T^2$ consistent and, thus, very precisely estimated. We say that the estimate of $\boldsymbol{\beta}$ is super-super consistent.

3. the tests of hypotheses on $\boldsymbol{\beta}$ are not tests of cointegration from $I(1)$ to $I(0)$, but instead from $I(2)$ to $I(1)$, as is evident from (14) and a cointegration relation should in general be considered $I(1)$, albeit noting that a cointegration relation $\boldsymbol{\beta}_i'\mathbf{x}_t$ can be CI(2,2), i.e. be cointegrating from $I(2)$ to $I(0)$,

4. the MA representation is essentially useless, as the once cumulated residuals cannot satisfactorily explain variables containing $I(2)$ trends, i.e. twice cumulated residuals.

Thus, one can test a number of hypotheses based on the $I(1)$ procedure even if $\mathbf{x}_t$ is $I(2)$, but the interpretation of the results has to be modified accordingly.

# 4 An $I(1)$ analysis of prices and exchange rate

## 4.1 Specification

The VAR model is based on the assumption of multivariate normality which, if correct, implies linearity in parameters as well as constancy of parameters. However, multivariate normality is seldom satisfied in a first tentatively estimated VAR model. There are many reasons for this, for example omission of relevant variables, inadequate measurements, interventions, reforms, etc. All this may have changed the data generating mechanisms, thus producing structural breaks, or resulting in extraordinary effects on some of the variables. In the present case, the reunification of East and West Germany in 1991:1 was a particularly important institutional event which is likely to have changed some of the properties of the VAR model. For example, Figure 1 shows that the nominal exchange rate may have experienced a change in its trending behavior at the reunification, as well as a shift in its level. Therefore, a consequence of merging the less productive East with West Germany is likely to have been a change in relative productivity, which needs to be accounted for by a change in the slopes of the linear trends in the VAR model.

Thus, in order to achieve a well-specified VAR model one usually has to control for major institutional events. Section 5.2 will provide a more detailed account of how to specify deterministic components in the $I(2)$ model. For the specification of such events in the $I(1)$ model the reader is referred to Juselius (2006, Chapter 6). Here they will be modelled by a trend with a changing slope at 1991:1 and various dummy variables, as explained below:

$$\Delta\mathbf{x}_t = \mathbf{\Gamma}_1\Delta\mathbf{x}_{t-1} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\mu}_0 + \boldsymbol{\mu}_{01}D_{s,91.1,t} + \boldsymbol{\mu}_1 t + \boldsymbol{\mu}_{11}t_{91.1} + \mathbf{\Phi}_p\mathbf{D}_{p,t} + \boldsymbol{\varepsilon}_t, \tag{15}$$

where the sample period is 1975:09-1998:12 and $\mathbf{x}_t' = [p_{1,t}, p_{2,t}, s_{12,t}]$ with

$p_{1,t}$ = log of German CPI[2],

$p_{2,t}$ = log of US CPI, and

$s_{12,t}$ = log of the nominal Dmk/$ exchange rate.

The linear terms in (15) are defined as

$\boldsymbol{\mu}_0$ is a vector of constant terms,

$\boldsymbol{\mu}_{01}$ is a vector measuring a change in the constant term at 1991:1

$\boldsymbol{\mu}_1$ a vector of linear trend slopes,

$\boldsymbol{\mu}_{11}$ a vector measuring a change in the trend slope at 1991:1.

The dummy variables are defined as:

$D_p tax = 1$ in 1991:7, 1991:9, and 1993:1, zero otherwise

$Ds91.1_t$ is 1 for $t \geq$ 1991:1, 0 otherwise,

$\mathbf{D}_{p,t}' = [Dp80.7, Dp91.1, D_p tax, D_p 97.7]$ with

$DpXX.y_t = 1$ in 19XX:y, zero otherwise.

The tax dummy is needed to account for a series of commodity tax increases to pay for reunification and the three dummies are needed to account for a big drop in the US inflation rate in 1980:7, the large changes in the nominal exchange rate in 1991:1 and 1997:7.

A*s discussed in more detail in section 8.6, the two trend components, the constant, and the shift dummy need to be appropriately restricted in the VAR model to avoid quadratic and cubic trends.* The dummy variables have been specified to exclusively control for the extraordinary shock at the time of the intervention, but to leave the information of the observation intact through its lagged impact. Thus, the dummies do not *remove* the outlying observation as is usually the case in a static regression model. Table 1 reports the estimated effects.

Conditional on the dummies, the VAR model becomes reasonably well-specified. The tests for multivariate residual autocorrelation at one lag, $\chi^2(9) = 11.0[0.28]$, and two lags, $\chi^2(9) = 14.2[0.12]$, were acceptable, as

[2] German CPI has been additively mean corrected for the reunification in 1991:1 prior to the VAR analysis.

Table 1: Estimated outlier effects and misspecification tests]

| Estimated outlier effects | | | | | Misspecification tests | | |
|---|---|---|---|---|---|---|---|
| | $D_p tax$ | $Dp80.7$ | $D_s 91.1$ | $D_p 97.7$ | $Norm.$ | $Skew.$ | $Kurt.$ |
| $\Delta p_{1,t}$ | **0.01** [11.36] | −0.00 [−1.40] | 0.00 [1.77] | **0.01** [4.15] | 7.22[0.03] | 0.35 | 3.62 |
| $\Delta p_{2,t}$ | −0.00 [−0.15] | **−0.01** [−4.90] | 0.00 [0.16] | 0.00 [0.37] | 15.4[0.00] | -0.20 | 4.20 |
| $\Delta s_{12,t}$ | −0.02 [−1.04] | 0.01 [0.39] | **0.01** [2.57] | **0.06** [1.98] | 6.31[0.04] | 0.10 | 3.66 |

$t$ - ratios in []

were the test of multivariate ARCH of order one, $\chi^2(36) = 45.9[0.12]$, and order two, $\chi^2(72) = 87.2[0.11]$. However, multivariate normality was rejected based on $\chi^2(6) = 27.1[0.00]$. To get some additional information, Table 1 reports the univariate Jarque-Bera tests, as well skewness (third moment around the mean) and kurtosis (forth moment around the mean). It appears that the non-normality problems are mostly due to excess kurtosis in the US inflation rate. Since the VAR estimates have been shown to be reasonably robust to moderate deviations from normality due to excess kurtosis (Gonzalo, 1994), the baseline VAR model is considered to be a reasonably adequate characterization of the data.

## 4.2 Rank determination and general model properties

The determination of the cointegration rank is a crucial step in the analysis, as it structures the data into its pulling and pushing components. The so called trace test (Johansen, 1988) is a likelihood ratio test for the cointegration rank. However, the trace test is derived under the null of $p - r$ unit roots, which does not always correpond to the null of the theory model as illustrated in Section 7 (see also Juselius, 2006, Chapter 8). Therefore, the choice of rank suggested by the trace test needs to be checked for its consistency with other information in the model, such as the characteristic roots.

The trace tests reported in Table 2 suggest a borderline acceptance of $r = 1$ cointegration relation and, hence, $p - r = 2$ common stochastic trends or, alternatively, a strong acceptance of $r = 2$ and, hence, $p - r = 1$ common stochastic trend. Thus, from a statistical point of view both choices can be defended. Section 7 will argue that $r = 2$ is the theory consistent choice. To find out which choice is econometrically preferable, we shall check the consistency of $r = 1, 2$ with the characteristic roots in the model and with the mean reversion of the cointegration relations.

Table 2: Determination of rank in the I(1) model

| $r$ | $p$-$r$ | $\boldsymbol{\tau}_{p-r}$ | 4 largest characteristic roots | | | |
|---|---|---|---|---|---|---|
| 0 | 3 | 80.06 [57.9] | 1.0 | 1.0 | 1.0 | 0.75 |
| 1 | 2 | **32.65** [36.6] | 1.0 | 1.0 | 0.99 | 0.53 |
| 2 | 1 | **6.72** [18.5] | 1.0 | 0.99 | 0.99 | 0.52 |
| 3 | 0 | | 0.99 | 0.99 | 0.98 | 0.53 |

Note: 95% quantiles in []

Tests of pushing and pulling variables

| | $r$ | $p_1$ | $p_2$ | $s_{12}$ |
|---|---|---|---|---|
| No levels feedback | 1 | 7.52 [0.01] | 16.17 [0.00] | 7.58 [0.01] |
| | 2 | 23.83 [0.00] | 32.74 [0.00] | 8.66 [0.01] |
| Pure adjustment | 1 | 21.40 [0.00] | 11.26 [0.00] | 34.27 [0.00] |
| | 2 | **2.74** [0.10] | **1.31** [0.25] | 18.74 [0.00] |

Note: p-values in []

An inspection of the characteristic roots of the model shows that there are three large roots of magnitude 0.99 in the unrestricted model. These are generally indistinguishable from unit roots, so the model seems to contain three unit roots. The choice of $r = 1$ leaves one near unit root and the choice of $r = 2$ two near unit roots in the model. Section 3 showed that, when one or several large roots remain in the model for any reasonable choice of $r$, it is a sign of $I(2)$ behavior in at least one of the variables.[3]

To check the consistency of the results with the $I(2)$ model, it is useful to divide the total number of stochastic trends into $I(1)$ and $I(2)$ trends, i.e. $p-r = s_1 + s_2$ where $s_1$ denotes the number of $I(1)$ trends (unit root processes), and $s_2$ the number of $I(2)$ trends (double unit root processes). Three (near) unit roots in the model would be consistent with either $\{r = 0, p - r = 3\}$ or $\{r = 1, s_1 = 1, s_2 = 1\}$, whereas $\{r = 2, s_1 = 0, s_2 = 1\}$ corresponds to two unit roots . Since the latter is less than the three near unit roots in the model, the choice $r = 2$ would not be consistent with the empirical information in the data.

Thus, by imposing $r = 1$, two of the big roots are restricted to unity,

[3]Note, however, that this diagnostic check is only reliable in a VAR model with a correct lag length. A VAR model with too many lags will often generate complex pairs of large (albeit insignificant) roots in the characteristic polynomial (Nielsen and Nielsen, 2006).

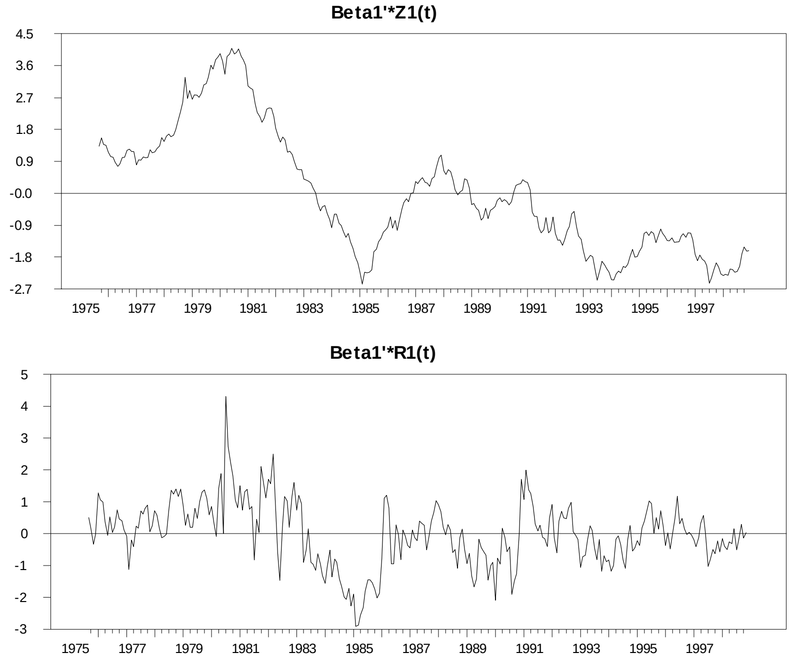


Figure 3: The graphs of the first cointegration relation. $\boldsymbol{\beta}'\mathbf{x}_t$ in the upper panel, $\boldsymbol{\beta}'\mathbf{R}_{1,t}$ in the lower panel.

but the third would still be unrestricted in the $I(1)$ model, invalidating some of the interpretation of the empirical results as discussed in Section 3. The graphs of the first two cointegration relations, reported in Figures 3 and 4, illustrate the effect of a near unit root. Based on the graphs, it is difficult to argue that $\boldsymbol{\beta}_i'\mathbf{x}_t$, $i = 1, 2$, is mean-reverting as an equilibrium error should be. However, $\boldsymbol{\beta}_i'\mathbf{R}_{1,t}$ (in the lower panel) looks much more mean-reverting, at least for $r = 1$. This, of course, is exactly in accordance with (13). Thus, only $\{r = 1, s_1 = 1, s_2 = 1\}$ seems acceptable based on the characteristic roots of the model and the graphs of the cointegration relations.

It is also useful to investigate the general pulling and pushing properties of the model described by the test of a unit vector in $\boldsymbol{\alpha}$ and a zero row in $\boldsymbol{\alpha}$ (Juselius, 2006, Chapter 11) and how they would be affected by the choice of rank. In the lower part of Table 2 the tests of 'no levels feed-back' (a zero row in $\boldsymbol{\alpha}$) and 'pure adjustment' (a unit vector in $\boldsymbol{\alpha}$) are reported for $r = 1$ and $r = 2$. For $r = 1$, none of the variables are found to purely pushing or pulling. For $r = 2$, there is some evidence that the two prices are exclusively adjusting (though the hypothesis that they are jointly adjusting is rejected).

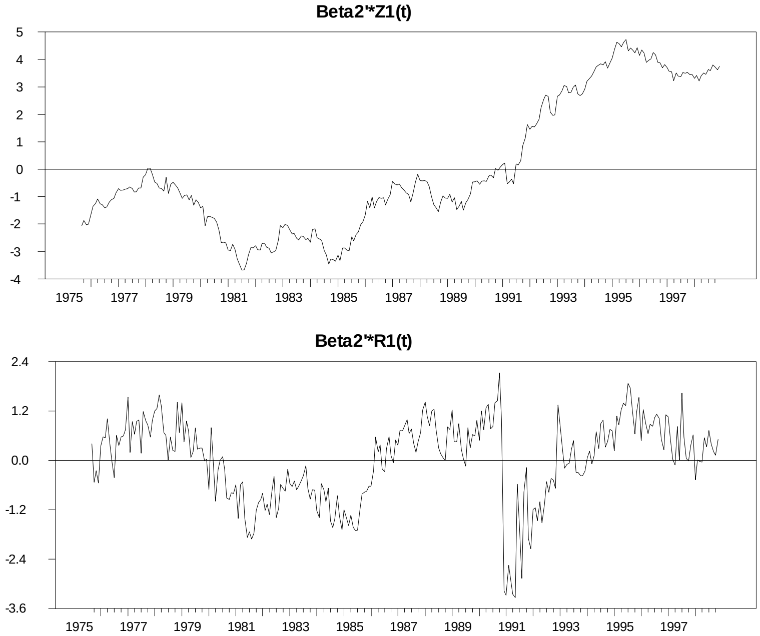


Figure 4: The graphs of the second cointegration relation. $\boldsymbol{\beta}'\mathbf{x}_t$ in the upper panel, $\boldsymbol{\beta}'\mathbf{R}_{1,t}$ in the lower panel.

Altogether, the empirical evidence suggests that prices are 'more' pulling than pushing which is an interesting observation as one would expect the opposite during a currency float.

## 4.3 Estimating the long-run structure

Table 3 reports the estimates of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Phi}$ for the choice of $r = 1$. The estimated $\boldsymbol{\beta}$ relation suggests that $p_{1,t}$ and $p_{2,t}$ are almost homogeneously related. Testing the hypothesis gives a test statistic $\chi^2(1) = 0.56[0.46]$ and, thus, price homogeneity of $\boldsymbol{\beta}'\mathbf{x}_t$ seems acceptable[4] when allowing for a broken trend. The presence of a broken linear trend might seem difficult to interpret but is probably a proxy for omitted variables effects, such as the effect of productivity differentials on relative prices, the so called Balassa-Samuelson effect (Balassa, 1964, Samuelson, 1964). The change in the trend slope at re-unification supports this interpretation. What is more surprising, however,

[4]When the data are I(2) price homogeneity of $\boldsymbol{\beta}'\mathbf{x}_t$ is a necessary, but not sufficient condition as will be discussed in Section 7.

Table 3: The estimated short-run dynamic adjustment structure

$$\underbrace{\begin{bmatrix} \Delta p_{1t} \\ \Delta p_{2,t} \\ \Delta s_{12,t} \end{bmatrix}}_{I(1)} = \underbrace{\begin{bmatrix} \underset{[4.50]}{\mathbf{0.21}} & \underset{[2.31]}{\mathbf{0.12}} & \underset{[2.06]}{\mathbf{0.01}} \\ \underset{[2.21]}{\mathbf{0.10}} & \underset{[10.23]}{\mathbf{0.52}} & \underset{[0.38]}{0.00} \\ \underset{[1.15]}{0.92} & \underset{[-1.59]}{-1.44} & \underset{[-0.18]}{-0.01} \end{bmatrix}}_{\Gamma_1} \underbrace{\begin{bmatrix} \Delta p_{1t-1} \\ \Delta p_{2,t-1} \\ \Delta s_{12,t-1} \end{bmatrix}}_{I(1)}$$

$$+ \underbrace{\begin{bmatrix} \underset{[-3.92]}{\mathbf{-0.01}} \\ \underset{[-5.92]}{\mathbf{-0.02}} \\ \underset{[-3.05]}{\mathbf{-0.17}} \end{bmatrix}}_{\boldsymbol{\alpha}} \underbrace{\begin{bmatrix} \boldsymbol{\beta}_1' \mathbf{x}_{t-1} \end{bmatrix}}_{I(1)} + \underbrace{\begin{bmatrix} \underset{[1.77]}{\mathbf{0.00}} & \underset{[4.09]}{\mathbf{0.02}} \\ \underset{[0.16]}{0.00} & \underset{[6.21]}{\mathbf{0.03}} \\ \underset{[2.57]}{\mathbf{0.01}} & \underset{[2.96]}{\mathbf{0.22}} \end{bmatrix}}_{\Phi} \begin{bmatrix} D_s 91.1 \\ \mu_0 \end{bmatrix} + \underbrace{\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}}_{I(0)}$$

where

$$\boldsymbol{\beta}_1' x_t = \mathbf{1.0} p_{1,t} \underset{[-7.76]}{\mathbf{-0.81}}\, p_{2,t} + \underset{[4.39]}{\mathbf{0.18}} s_{12,t} \underset{[-4.81]}{\mathbf{-0.0022}}\, t_{91.1} + \underset{[3.96]}{\mathbf{0.0022}} t$$

and

$$\Omega = \begin{bmatrix} 1.00 & & \\ 0.12 & 1.00 & \\ -0.58 & -0.07 & 1.00 \end{bmatrix}$$

is that the sign of the nominal exchange rate is opposite to the expected one. Based on Figure 1 it is easy to see why: over the sample period relative prices and nominal exchange rates have frequently moved in opposite directions for extended periods of time. For this reason, the data do not support the *ppp* restrictions (1,-1,-1) on $\boldsymbol{\beta}$.

The estimated $\boldsymbol{\alpha}$ coefficients show that German prices and nominal exchange rates have been equilibrium correcting to the estimated $\boldsymbol{\beta}$ relation whereas US prices have been increasing in the equilibrium errors. The overall behavior of the system is, nevertheless, stable as the other two variables compensate for the error increasing behavior of US prices. The estimated coefficients of $\mathbf{\Gamma}_1$ show that lagged inflation rates are quite significant in the price equations, whereas the lagged depreciation/appreciation rate is only significant in the German price equation. As already demonstrated in Section 3, the lagged changes of the $I(2)$ variables in $\mathbf{\Gamma}_1$ are needed to achieve stationarity of $\boldsymbol{\beta}_1'\mathbf{R}_{1,t}$.

The estimates of $\boldsymbol{\alpha}_{\perp 1,}$ $\boldsymbol{\beta}_{\perp 1}$ and $\mathbf{C}$ in the MA representation of the $I(1)$ model are almost all insignificant and are not reported here. This is because the stochastic trends in the $I(1)$ model are measured by the once cumulated residuals, whereas the data are generated by second order stochastic trends, measured by the twice cumulated residuals. Thus, when data are $I(2)$ the MA representation of the $I(1)$ model is completely uninformative.

Based on the above results, it would be hard to argue that the data are not empirically $I(2)$ and the next step is, therefore, to address the PPP puzzle in the correct framework of an $I(2)$ model.

# 5 Representing the $I(2)$ model

## 5.1 The basic structure

As discussed in Section 1, formulation (3) is convenient when data are $I(2)$:

$$\Delta^2 \mathbf{x}_t = \mathbf{\Gamma \Delta x_{t-1}} + \mathbf{\Pi x_{t-1}} + \boldsymbol{\mu}_\mathbf{0} + \boldsymbol{\mu}_\mathbf{01} D_{s,91.1,t} + \boldsymbol{\mu}_\mathbf{1} t + \boldsymbol{\mu}_\mathbf{11} t_{91.1} + \mathbf{\Phi}_p \mathbf{D}_{p,t} + \boldsymbol{\varepsilon}_t, \tag{16}$$

where the deterministic components are in Section 4.1. Similar to the $I(1)$ model, we need to define the concentrated $I(2)$ model[5]:

$$\mathbf{R}_{0,t} = \mathbf{\Gamma R}_{1,t} + \mathbf{\Pi R}_{2,t} + \boldsymbol{\varepsilon}_t \tag{17}$$

where $\mathbf{R}_{0,t}$, $\mathbf{R}_{1,t}$, and $\mathbf{R}_{2,t}$ are defined by:

$$\Delta^2 \mathbf{x}_t = \hat{\mathbf{b}}_{10} + \hat{\mathbf{b}}_{11} t + \hat{\mathbf{B}}_{11} \mathbf{D}_{s,t} + \hat{\mathbf{B}}_{12} \mathbf{D}_{p,t} + \mathbf{R}_{0,t}, \tag{18}$$

$$\Delta \tilde{\mathbf{x}}_{t-1} = \hat{\mathbf{b}}_{20} + \hat{\mathbf{b}}_{21} t + \hat{\mathbf{B}}_{21} \mathbf{D}_{s,t} + \hat{\mathbf{B}}_{22} \mathbf{D}_{p,t} + \mathbf{R}_{1,t}, \tag{19}$$

$$\tilde{\mathbf{x}}_{t-1} = \hat{\mathbf{b}}_{30} + \hat{\mathbf{b}}_{31} t + \hat{\mathbf{B}}_{31} \mathbf{D}_{s,t} + \hat{\mathbf{B}}_{32} \mathbf{D}_{p,t} + \mathbf{R}_{2,t}. \tag{20}$$

and $\tilde{\mathbf{x}}_t$ indicates that $\mathbf{x}_t$ has been augmented with some deterministic components such as trend, constant, and shift dummy variables. The matrices $\mathbf{\Pi}$ and $\mathbf{\Gamma}$ are subject to the two reduced rank restrictions, $\mathbf{\Pi} = \boldsymbol{\alpha}'\boldsymbol{\beta}$, where $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are $p \times r$, and $\boldsymbol{\alpha}'_\perp \mathbf{\Gamma} \boldsymbol{\beta}_\perp = \boldsymbol{\xi}\boldsymbol{\eta}'$, where $\boldsymbol{\xi}, \boldsymbol{\eta}$ are $(p-r) \times s_1$. The model in (16) contains an unrestricted constant with a shift, a broken trend and a few impulse dummies that will have to be adequately restricted to avoid undesirable effects, as discussed in Section 5.2.

The moving average representation of the $I(2)$ model (Johansen, 1992b, 1996,1997) with unrestricted deterministic components is given by:

[5] When the lag $k > 2$, there would also be lagged acceleration rates, $\Delta^2 \mathbf{x}_{t-1}$, to concentrate out.

$$
\begin{aligned}
\mathbf{x}_t &= \mathbf{C}_2 \sum_{j=1}^{t} \sum_{i=1}^{j} (\boldsymbol{\varepsilon}_i + \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 i + \boldsymbol{\mu}_{01} D_{s,91.1,i} + \boldsymbol{\Phi}_p \mathbf{D}_{p,i}) \\
&\quad + \mathbf{C}_1 \sum_{j=1}^{t} (\boldsymbol{\varepsilon}_j + \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 j + \boldsymbol{\mu}_{01} D_{s,91.1,j} + \boldsymbol{\Phi}_p \mathbf{D}_{p,j}) \\
&\quad + \mathbf{C}^*(L)(\boldsymbol{\varepsilon}_t + \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\mu}_{01} D_{s,91.1,t} + \boldsymbol{\Phi}_p \mathbf{D}_{p,t}) + \mathbf{A} + \mathbf{B}t
\end{aligned} \tag{21}
$$

where $\mathbf{A}$ and $\mathbf{B}$ are functions of the initial values $\mathbf{x}_0, \mathbf{x}_{-1}, \mathbf{x}_{-2}$, and the coefficient matrices satisfy:

$$\mathbf{C}_2 = \boldsymbol{\beta}_{\perp 2}(\boldsymbol{\alpha}'_{\perp 2}\Psi\boldsymbol{\beta}_{\perp 2})^{-1}\boldsymbol{\alpha}'_{\perp 2},$$

$$
\begin{aligned}
\boldsymbol{\beta}'\mathbf{C}_1 &= -\overline{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}\mathbf{C}_2, \quad \boldsymbol{\beta}'_{\perp 1}\mathbf{C}_1 = -\overline{\boldsymbol{\alpha}}'_{\perp 1}(\mathbf{I} - \boldsymbol{\Psi}\mathbf{C}_2), \\
\boldsymbol{\Psi} &= \boldsymbol{\Gamma}\overline{\boldsymbol{\beta}}\overline{\boldsymbol{\alpha}}'\boldsymbol{\Gamma} + \mathbf{I} - \boldsymbol{\Gamma}_1
\end{aligned} \tag{22}
$$

where the notation $\overline{\boldsymbol{\alpha}} = \boldsymbol{\alpha}(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}$ is used all through the chapter. To facilitate the interpretation of the $I(2)$ stochastic trends and how they load into the variables, it is useful to let $\tilde{\boldsymbol{\beta}}_{\perp 2} = \boldsymbol{\beta}_{\perp 2}(\boldsymbol{\alpha}'_{\perp 2}\boldsymbol{\Psi}\boldsymbol{\beta}_{\perp 2})^{-1}$, so that

$$\mathbf{C}_2 = \tilde{\boldsymbol{\beta}}_{\perp 2}\boldsymbol{\alpha}'_{\perp 2}. \tag{23}$$

It is now easy to see that the $\mathbf{C}_2$ matrix has a similar reduced rank representation to $\mathbf{C}_1$ in the $I(1)$ model, so it is straightforward to interpret $\boldsymbol{\alpha}'_{\perp 2}\sum\sum\boldsymbol{\varepsilon}_i$ as a measure of the $s_2$ second order stochastic trends which load into the variables $\mathbf{x}_t$ with the weights $\tilde{\boldsymbol{\beta}}_{\perp 2}$.

From (22) we note that the $\mathbf{C}_1$ matrix in the $I(2)$ model cannot be given a simple decomposition as it depends on both the $\mathbf{C}_2$ matrix and the other model parameters in a complex way. Johansen (2007) derives an analytical expression for $\mathbf{C}_1$, essentially showing that:

$$\mathbf{C}_1 = \boldsymbol{\omega}_0\boldsymbol{\alpha}' + \boldsymbol{\omega}_1\boldsymbol{\alpha}'_{\perp 1} + \boldsymbol{\omega}_2\boldsymbol{\alpha}'_{\perp 2} \tag{24}$$

where $\boldsymbol{\omega}_i$ are complicated functions of the parameters of the model (not to be reproduced here).

To summarize the basic structures of the $I(2)$ model, Table 4 decomposes the vector $\mathbf{x}_t$ into the directions of $(\boldsymbol{\beta}, \boldsymbol{\beta}_{\perp 1}, \boldsymbol{\beta}_{\perp 2})$ and the directions of $(\boldsymbol{\alpha}, \boldsymbol{\alpha}_{\perp 1}, \boldsymbol{\alpha}_{\perp 2})$. The left hand side of the table illustrates the $\boldsymbol{\beta}, \boldsymbol{\beta}_{\perp}$ directions, where $\boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\delta}'\Delta\mathbf{x}_t$ defines the stationary polynomially cointegrating relation, and $\boldsymbol{\beta}'_{\perp 1}\mathbf{x}_t$ the $CI(2,1)$ relation that can only become stationary by differencing. The $\boldsymbol{\beta}, \boldsymbol{\beta}_{\perp 1}$ relations define the two stationary cointegration relations between the differenced variables, $\boldsymbol{\tau}'\Delta\mathbf{x}_t$. Finally, $\boldsymbol{\beta}'_{\perp 2}\mathbf{x}_t \sim I(2)$ is a

Table 4: Decomposing the data vector using the I(2) model

| | The $\boldsymbol{\beta}, \boldsymbol{\beta}_{\perp}$ decomposition of $\mathbf{x}_t$ | The $\boldsymbol{\alpha}, \boldsymbol{\alpha}_{\perp}$ decomposition |
|---|---|---|
| $r = 1$ | $[\underbrace{\boldsymbol{\beta}_1' \mathbf{x}_t}_{I(1)} + \underbrace{\boldsymbol{\delta}_1' \Delta \mathbf{x}_t}_{I(1)}] \sim I(0)$ | $\boldsymbol{\alpha}_1$: short-run adjustment coefficients |
| $s_1 = 1$ | $\boldsymbol{\beta}_{\perp 1}' \mathbf{x}_t \sim I(1)$ | $\boldsymbol{\alpha}_{\perp 1}' \sum_{i=1}^{t} \boldsymbol{\varepsilon}_i$: $I(1)$ stochastic trend |
| $p - s_2 = 2$ | $\boldsymbol{\tau}' \Delta \mathbf{x}_t = (\boldsymbol{\beta}, \boldsymbol{\beta}_{\perp 1})' \Delta \mathbf{x}_t \sim I(0)$ | |
| $s_2 = 1$ | $\boldsymbol{\beta}_{\perp 2}' \mathbf{x}_t = \boldsymbol{\tau}_{\perp}' \mathbf{x}_t \sim I(2)$ | $\boldsymbol{\alpha}_{\perp 2}' \sum_{s=1}^{t} \sum_{i=1}^{s} \boldsymbol{\varepsilon}_i$: $I(2)$ stochastic trend |

non-cointegrating relation, which can only become stationary by differencing twice. The right hand side of the table illustrates the corresponding decomposition into the $\boldsymbol{\alpha}, \boldsymbol{\alpha}_{\perp}$ directions, where $\boldsymbol{\alpha}$ define the dynamic adjustment coefficients to the polynomially cointegrating relation, whereas $\boldsymbol{\alpha}_{\perp 1}$ and $\boldsymbol{\alpha}_{\perp 2}$ define the first and second order stochastic trends as a linear function of the VAR residuals.

## 5.2 Deterministic components

A correct specification of the deterministic components, such as trends, constant and dummies, and how they enter the model is mandatory for the $I(2)$ analysis. This is because the chosen specification is likely to strongly affect the reliability of the model estimates and to change the asymptotic distribution of the rank test. Because the typical smooth behavior of a stochastic $I(2)$ trend sometimes can be approximated with an $I(1)$ stochastic trend around a broken linear deterministic trend, one can in some cases avoid the $I(2)$ analysis altogether by allowing for sufficiently many breaks in the linear trend. Whether one specification is preferable to the other is difficult to know, but we need to pay sufficient attention to this question, as the choice is likely to significantly influence the empirical results.

In the present data, the reunification of Germany is likely to have significantly affected German prices, but not US prices. The raw data exhibit an extraordinary large shock in $\Delta^2 p_{1,t}$ due to the reunification in 1991:1. A big impulse in $\Delta^2 p_{1,t}$ cumulates to a level shift in $\Delta p_{1,t}$, and double cumulates to a broken linear trend in $p_{1,t}$. Thus, accounting for the extraordinary large shock at 1991:1 with a blip dummy in $\Delta^2 p_{1,t}$, a shift dummy in $\Delta p_{1,t}$ is econometrically consistent with broken linear trends in prices. Because such a broken linear trend may or may not cancel in $\boldsymbol{\beta}'\mathbf{x}_t$, the model should be specified to allow for a (testable) broken linear trend in $\boldsymbol{\beta}'\mathbf{x}_t$. Likewise, the level shift may (or may not) cancel in $\boldsymbol{\delta}'\Delta\mathbf{x}_t$ or $\boldsymbol{\tau}'\Delta\mathbf{x}_t$. Thus, the model specification should allow for this possibility. Inspecting the graphs in Figure 1 shows an increasing trend in both prices and a downward sloping trend in relative prices and the question is whether the latter is cancelled by cointegration with the nominal exchange rate.

Whatever the case, quadratic or cubic trends will be excluded from the outset and the model specification should account for this.

To understand the role of the deterministic terms in the $I(2)$ model, it is useful to specify the mean of the stationary parts of (16) allowing for the above effects (so that they can be tested), while at the same time excluding cubic or quadratic trend effects.

The mean of $\Delta^2\mathbf{x}_t$ should be allowed to contain the impulse dummies as these do not double cumulate to quadratic trends, i.e.:

$$E\Delta^2\mathbf{x}_t = \boldsymbol{\Phi}_p\mathbf{D}_{p,t}$$

The mean of the polynomially cointegrated relations should be allowed to have a trend and a broken linear trend in $\boldsymbol{\beta}'\mathbf{x}_t$ and a constant and a shift dummy in $\boldsymbol{\delta}'\Delta\mathbf{x}_t$, i.e.:

$$E(\boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\delta}'\Delta\mathbf{x}_t) = \boldsymbol{\rho}_0 t + \boldsymbol{\rho}_{01} t_{91.1} + \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}_{01} D_s 91.1_t \tag{25}$$

The mean of the difference stationary relations $\boldsymbol{\tau}'\Delta\mathbf{x}_t$ should be allowed to contain a step dummy and a constant, i.e.:

$$E(\boldsymbol{\tau}'\Delta\mathbf{x}_t) = \boldsymbol{\omega}_0 + \boldsymbol{\omega}_{01} D_s 91.1_t$$

The question is now how to restrict $\boldsymbol{\mu}_0, \boldsymbol{\mu}_{01}$, $\boldsymbol{\mu}_1$, and $\boldsymbol{\mu}_{11}$ in (16)[6] to allow for the deterministic components in the above mean values while suppressing any quadratic or cubic trend effects in the model. The general idea will only be demonstrated for the constant term $\boldsymbol{\mu}_0$ and the linear term $\boldsymbol{\mu}_1$ as the procedure is easily generalized to the step dummy and the broken trend. A more detailed discussion is given in Juselius (2006, Chapter 17).

First, the constant term $\boldsymbol{\mu}_0$ is decomposed into three components proportional to $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}_{\perp 1}$ and $\boldsymbol{\alpha}_{\perp 2}$:

$$\boldsymbol{\mu}_0 = \boldsymbol{\alpha}\boldsymbol{\gamma}_0 + \boldsymbol{\alpha}_{\perp 1}\boldsymbol{\gamma}_1 + \boldsymbol{\alpha}_{\perp 2}\boldsymbol{\gamma}_2. \tag{26}$$

The step dummy $\boldsymbol{\mu}_{01}$ is similarly decomposed:

[6] At this stage, $\boldsymbol{\Phi}_p$ will be left unrestriced in the model.

$$\boldsymbol{\mu}_{01} = \boldsymbol{\alpha}\boldsymbol{\gamma}_{01} + \boldsymbol{\alpha}_{\perp 1}\boldsymbol{\gamma}_{11} + \boldsymbol{\alpha}_{\perp 2}\boldsymbol{\gamma}_{21}$$

To investigate the effect of an unrestricted constant on $\mathbf{x}_t$, (26) is then inserted in (21) using (23) and (24) . The effect of cumulating the constant twice is given by:

$$\begin{aligned}
\mathbf{C}_2 \sum_{j=1}^{t}\sum_{i=1}^{j} \boldsymbol{\mu}_0 &= \sum_{j=1}^{t}\sum_{i=1}^{j} \tilde{\boldsymbol{\beta}}_{\perp 2}\boldsymbol{\alpha}'_{\perp 2}(\boldsymbol{\alpha}\boldsymbol{\gamma}_0 + \boldsymbol{\alpha}_{\perp 1}\boldsymbol{\gamma}_1 + \boldsymbol{\alpha}_{\perp 2}\boldsymbol{\gamma}_2) \\
&= \tilde{\boldsymbol{\beta}}_{\perp 2}\boldsymbol{\alpha}'_{\perp 2}\boldsymbol{\alpha}_{\perp 2}\boldsymbol{\gamma}_2(t(t-1)/2) \qquad (27)
\end{aligned}$$

as $\boldsymbol{\alpha}'_{\perp 2}\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\alpha}'_{\perp 2}\boldsymbol{\alpha}_{\perp 1} = \mathbf{0}$. Thus, an *unrestricted constant* term in the VAR model will allow for a quadratic trend in $\mathbf{x}_t$ so we need to restrict the $\boldsymbol{\alpha}_{\perp 2}$ component of $\boldsymbol{\mu}_0$ to avoid this. How to do it will be discussed below.

The effect of cumulating the constant term once is given by:

$$\begin{aligned}
\mathbf{C}_1 \sum_{j=1}^{t} \boldsymbol{\mu}_0 &= (\boldsymbol{\omega}_0\boldsymbol{\alpha}' + \boldsymbol{\omega}_1\boldsymbol{\alpha}'_{\perp 1} + \boldsymbol{\omega}_2\boldsymbol{\alpha}'_{\perp 2})\sum_{j=1}^{t}(\boldsymbol{\alpha}\boldsymbol{\gamma}_0 + \boldsymbol{\alpha}_{\perp 1}\boldsymbol{\gamma}_1 + \boldsymbol{\alpha}_{\perp 2}\boldsymbol{\gamma}_2) \\
&= [(\boldsymbol{\omega}_0\underbrace{\boldsymbol{\alpha}'\boldsymbol{\alpha}\boldsymbol{\gamma}_0}_{\widetilde{\gamma_0}} + \boldsymbol{\omega}_1\underbrace{\boldsymbol{\alpha}'_{\perp 1}\boldsymbol{\alpha}_{\perp 1}\boldsymbol{\gamma}_1}_{\widetilde{\gamma_1}} + \boldsymbol{\omega}_2\underbrace{\boldsymbol{\alpha}'_{\perp 2}\boldsymbol{\alpha}_{\perp 2}\boldsymbol{\gamma}_2}_{\widetilde{\gamma_2}})]t \qquad (28)
\end{aligned}$$

as $\boldsymbol{\alpha}'\boldsymbol{\alpha}_{\perp 1}= \mathbf{0}$, $\boldsymbol{\alpha}'\boldsymbol{\alpha}_{\perp 2}= \mathbf{0}$ and $\boldsymbol{\alpha}'_{\perp 1}\boldsymbol{\alpha}_{\perp 2} = \mathbf{0}$. Thus, there are three different linear trends associated with the $\mathbf{C}_1$ components of the constant term.

Most applications of the $I(2)$ model are for nominal variables implying that linear trends in the data is a natural starting hypothesis (as average nominal growth rates are generally nonzero). To achieve similarity in the rank test procedure (Nielsen and Rahbek, 2000), the model should allow for linear trends in all directions consistent with the specification of trend-stationarity as a starting hypothesis in (25). This means that $\boldsymbol{\mu}_1 t \neq \mathbf{0}$ and $\boldsymbol{\mu}_{11}t_{91.1} \neq \mathbf{0}$ in (16), so the vector $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_{11}$ need to be decomposed similarly to the constant term and the step dummy:

$$\boldsymbol{\mu}_1 = \boldsymbol{\alpha}\boldsymbol{\rho}_0 + \boldsymbol{\alpha}_{\perp 1}\boldsymbol{\rho}_1 + \boldsymbol{\alpha}_{\perp 2}\boldsymbol{\rho}_2$$

and

$$\boldsymbol{\mu}_{11} = \boldsymbol{\alpha}\boldsymbol{\rho}_{01} + \boldsymbol{\alpha}_{\perp 1}\boldsymbol{\rho}_{11} + \boldsymbol{\alpha}_{\perp 2}\boldsymbol{\rho}_{21}.$$

We now focus on the linear trend term. The effect of cumulating this term twice is given by:

$$\begin{aligned}
\mathbf{C}_2 \sum_{j=1}^{t} \sum_{i=1}^{j} \boldsymbol{\mu}_1 i &= \sum_{j=1}^{t} \sum_{i=1}^{j} \boldsymbol{\beta}_{\perp 2} \boldsymbol{\alpha}'_{\perp 2} (\boldsymbol{\alpha} \boldsymbol{\rho}_0 + \boldsymbol{\alpha}_{\perp 1} \boldsymbol{\rho}_1 + \boldsymbol{\alpha}_{\perp 2} \boldsymbol{\rho}_2) i \\
&= \sum_{j=1}^{t} \sum_{i=1}^{j} \boldsymbol{\beta}_{\perp 2} \boldsymbol{\alpha}'_{\perp 2} \underbrace{\boldsymbol{\alpha}_{\perp 2} \boldsymbol{\rho}_2}_{=0} i \qquad (29)
\end{aligned}$$

Thus, unless we restrict $\boldsymbol{\alpha}_{\perp 2} \boldsymbol{\rho}_2 = \mathbf{0}$ the model will allow for cubic trends in the data. The $I(2)$ procedure in CATS in RATS (Dennis et al. 2005) imposes this restriction. The effect of cumulating the linear trend term once is given by:

$$\begin{aligned}
\mathbf{C}_1 \sum_{j=1}^{t} \boldsymbol{\mu}_1 j &= \sum_{j=1}^{t} (\boldsymbol{\omega}_0 \boldsymbol{\alpha}' + \boldsymbol{\omega}_1 \boldsymbol{\alpha}'_{\perp 1} + \boldsymbol{\omega}_2 \boldsymbol{\alpha}'_{\perp 2})(\boldsymbol{\alpha} \boldsymbol{\rho}_0 + \boldsymbol{\alpha}_{\perp 1} \boldsymbol{\rho}_1 + \boldsymbol{\alpha}_{\perp 2} \boldsymbol{\rho}_2) j \\
&= \sum_{j=1}^{t} (\boldsymbol{\omega}_0 \boldsymbol{\alpha}' \underbrace{\boldsymbol{\alpha} \boldsymbol{\rho}_0}_{\neq 0} + \boldsymbol{\omega}_1 \boldsymbol{\alpha}'_{\perp 1} \underbrace{\boldsymbol{\alpha}_{\perp 1} \boldsymbol{\rho}_1}_{=0} + \boldsymbol{\omega}_2 \boldsymbol{\alpha}'_{\perp 2} \underbrace{\boldsymbol{\alpha}_{\perp 2} \boldsymbol{\rho}_2}_{=0}) j \qquad (30)
\end{aligned}$$

It appears that all three $\mathbf{C}_1$ components of the linear trend will generate quadratic trends in the data. Based on (29) we already know that $\boldsymbol{\alpha}_{\perp 2} \boldsymbol{\rho}_2 = \mathbf{0}$. Unless we are willing to accept linear trends in $\boldsymbol{\alpha}'_{\perp 1} \Delta \mathbf{x}_t$[7], we should also restrict $\boldsymbol{\alpha}_{\perp 1} \boldsymbol{\rho}_1 = \mathbf{0}$. This leaves us with the $\boldsymbol{\alpha}$ component of $\mathbf{C}_1$, which cannot be set to zero, because $\boldsymbol{\alpha} \boldsymbol{\rho}_0 \neq \mathbf{0}$ is needed to allow for a linear trend in $\boldsymbol{\beta}' \mathbf{x}_t$. The problem is that a linear trend in a polynomially cointegrating relation, unless adequately restricted, generates a quadratic trend in $\mathbf{x}_t$. However, this can be solved by noticing that $\boldsymbol{\alpha}_{\perp 2} \gamma_2 \neq \mathbf{0}$ in (27) also generates a quadratic trend in $\mathbf{x}_t$, so that by restricting $\boldsymbol{\omega}_0 \boldsymbol{\alpha}' \boldsymbol{\alpha} \boldsymbol{\rho}_0 = -\boldsymbol{\beta}_{\perp 2} \boldsymbol{\alpha}'_{\perp 2} \boldsymbol{\alpha}_{\perp 2} \gamma_2$, the two trend components cancel and there will be no quadratic trends in the data. The trend-stationary polynomially cointegrated relation in Rahbek et al. (1999) was estimated subject to this constraint.

To summarize: To avoid quadratic and cubic trends in the $I(2)$ model we need to impose the following restrictions: $\boldsymbol{\rho}_1 = \boldsymbol{\rho}_2 = \mathbf{0}$ and $\boldsymbol{\omega}_0 \boldsymbol{\alpha}' \boldsymbol{\alpha} \boldsymbol{\rho}_0 = -\boldsymbol{\beta}_{\perp 2} \boldsymbol{\alpha}'_{\perp 2} \boldsymbol{\alpha}_{\perp 2} \gamma_2$, as well as $\boldsymbol{\rho}_{11} = \boldsymbol{\rho}_{21} = \mathbf{0}$ and $\boldsymbol{\omega}_0 \boldsymbol{\alpha}' \boldsymbol{\alpha} \boldsymbol{\rho}_{01} = -\boldsymbol{\beta}_{\perp 2} \boldsymbol{\alpha}'_{\perp 2} \boldsymbol{\alpha}_{\perp 2} \gamma_{21}$ to avoid broken quadratic and cubic trends.

[7] A linear trend in $\boldsymbol{\alpha}'_{\perp 1} \Delta \mathbf{x}_t$ would imply that inflation rate, say, is allowed to grow with a linear trend and, thus, prices with a quadratic trend. It would be hard to argue for such a specification except, possibly, as a local approximation.

# 6 Estimation in the $I(2)$ model

Johansen (1992) provided the solution to the two step estimator and Johansen (1997) to the full ML estimatior. Even though the two-stage procedure gives asymptotically efficient ML estimates, the small sample properties of the ML estimates are generally superior (Nielsen and Rahbek, 2007) and all subsequent results are based on the ML procedure.

## 6.1 The ML procedure

Section 1 showed that there is an important difference between the first and second rank condition. The former is formulated as a reduced rank condition directly on $\mathbf{\Pi}$, whereas the latter is on a transformed $\mathbf{\Gamma}$. The full ML procedure exploits the fact that the $I(2)$ model contains $p-s_2$ cointegration relations, $\boldsymbol{\tau}'\mathbf{x}_t$, where $\boldsymbol{\tau}=(\boldsymbol{\beta},\boldsymbol{\beta}_{\perp 1})$ define $r+s_1=p-s_2$ directions in which the process is cointegrated from $I(2)$ to $I(1)$. This means that $\boldsymbol{\tau}$ can be determined by solving just one reduced rank regression, after which the vector space can be divided into $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{\perp 1}$. This is the basic reason for the following parameterization of the I(2) model proposed by Johansen (1997):

$$\begin{aligned}\underbrace{\Delta^2\mathbf{x}_t}_{I(0)} &= \boldsymbol{\alpha}\underbrace{(\boldsymbol{\rho}'\underbrace{\tilde{\boldsymbol{\tau}}'\tilde{\mathbf{x}}_{t-1}}_{I(1)}+\underbrace{\tilde{\boldsymbol{\psi}}'\Delta\tilde{\mathbf{x}}_{t-1}}_{I(1)})}_{I(0)}+\boldsymbol{\omega}'\underbrace{\tilde{\boldsymbol{\tau}}'\Delta\tilde{\mathbf{x}}_{t-1}}_{I(0)}+\mathbf{\Phi}_p\mathbf{D}_{p,t}+\mathbf{\Phi}_{tr}\mathbf{D}_{tr,t}+\boldsymbol{\varepsilon}_t,\\ \boldsymbol{\varepsilon}_t &\sim N_p(\mathbf{0},\mathbf{\Omega}\ ),\ t=1,...,T\end{aligned} \tag{31}$$

where $\boldsymbol{\rho}=(\mathbf{I},\mathbf{0})$ is a $(r+s_1)\times r$ selection matrix designed to pick out the $r$ cointegration vectors $\boldsymbol{\beta}'\mathbf{x}_t$ (so that $\boldsymbol{\rho}'\boldsymbol{\tau}'=\boldsymbol{\beta}'$), $\boldsymbol{\psi}'=-(\boldsymbol{\alpha}'\mathbf{\Omega}^{-1}\boldsymbol{\alpha}\ )^{-1}\boldsymbol{\alpha}'\mathbf{\Omega}^{-1}\mathbf{\Gamma}$, $\boldsymbol{\omega}'=-\mathbf{\Omega}\boldsymbol{\alpha}_{\perp}(\boldsymbol{\alpha}_{\perp}'\mathbf{\Omega}\boldsymbol{\alpha}_{\perp})^{-1}(\boldsymbol{\alpha}_{\perp}'\mathbf{\Gamma}\bar{\boldsymbol{\beta}},\boldsymbol{\xi})$, $\boldsymbol{\rho}'\tilde{\boldsymbol{\tau}}'=[\boldsymbol{\beta}',\boldsymbol{\rho}_0,\boldsymbol{\rho}_{01}]$, $\tilde{\boldsymbol{\psi}}'=[\boldsymbol{\psi}',\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_{01}]$, $\tilde{\mathbf{x}}_t'=[\mathbf{x}_t',t,t83]$ and $\Delta\tilde{\mathbf{x}}_t'=[\Delta\mathbf{x}_t',1,Ds831]$.

The $FIML$ estimates of $\boldsymbol{\tau}=(\boldsymbol{\beta},\boldsymbol{\beta}_{\perp 1})$ are obtained by an iterative procedure which at each step delivers the solution of just one reduced rank problem and the eigenvectors give the estimates of $\boldsymbol{\tau}$. Thus, the vector $\mathbf{x}_t$ is decomposed into the $p-s_2$ directions $\boldsymbol{\tau}=(\boldsymbol{\beta},\boldsymbol{\beta}_{\perp 1})$ in which the process is $I(1)$ and the $s_2$ directions $\boldsymbol{\tau}_{\perp}=\boldsymbol{\beta}_{\perp 2}$ in which it is $I(2)$. For given values of $(\boldsymbol{\beta},\boldsymbol{\beta}_{\perp 1})$ it is possible to derive all the remaining matrices $(\boldsymbol{\alpha},\boldsymbol{\alpha}_{\perp 1},\boldsymbol{\alpha}_{\perp 2},\boldsymbol{\beta}_{\perp 2})$.

The matrix $\boldsymbol{\psi}$ in (31) does not make a distinction between stationary and nonstationary components in $\Delta\mathbf{x}_t$. For example, when $\mathbf{x}_t$ contains variables which are $I(2)$, for example prices, as well as $I(1)$, for example nominal exchange rates, then some of the differenced variables picked up by $\boldsymbol{\psi}$ will be $I(0)$. As the latter do not contain any stochastic $I(1)$ trends, they are by definition redundant in the polynomially cointegrated relations. The idea behind the parameterization in Paruolo and Rahbek (1997) was to express the polynomially cointegrated relations exclusively in terms of the differences of the $I(2)$ variables. This was achieved by noticing that

$$\boldsymbol{\psi}'\Delta\mathbf{x}_{t-1} = \boldsymbol{\psi}'(\boldsymbol{\tau}\boldsymbol{\tau}' + \boldsymbol{\tau}_{\perp}\boldsymbol{\tau}_{\perp}')\Delta\mathbf{x}_{t-1}.$$

The model given below is based on the Paruolo and Rahbek parameterization. As discussed in Section 5, the (broken) trend has been restricted to be proportional to $\boldsymbol{\alpha}$, and the constant and the shift dummy to be proportional to $\boldsymbol{\zeta}$.

$$\begin{aligned}\underbrace{\Delta^2\mathbf{x}_t}_{I(0)} &= \boldsymbol{\alpha}\underbrace{\left\{\underbrace{[\boldsymbol{\beta}',\boldsymbol{\rho}_0,\boldsymbol{\rho}_{01}]\begin{bmatrix}\mathbf{x}_{t-1}\\ t\\ t_{91.1}\end{bmatrix}}_{I(1)} + \underbrace{[\boldsymbol{\delta}',\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_{01}]\begin{bmatrix}\Delta\mathbf{x}_{t-1}\\ c\\ D_s91.1_{t-1}\end{bmatrix}}_{I(1)}\right\}}_{I(0)} \\ &+\boldsymbol{\zeta}\underbrace{\begin{bmatrix}\underbrace{\begin{matrix}\boldsymbol{\beta}',\boldsymbol{\rho}_0,\boldsymbol{\rho}_{01}\\ \boldsymbol{\beta}_{\perp1}',\tilde{\boldsymbol{\gamma}}_0,\tilde{\boldsymbol{\gamma}}_{01}\end{matrix}}_{\boldsymbol{\tau}}\end{bmatrix}\begin{bmatrix}\Delta\mathbf{x}_{t-1}\\ c\\ D_s91.1_{t-1}\end{bmatrix}}_{I(0)} + \boldsymbol{\Phi}_p\mathbf{D}_{p,t} + \boldsymbol{\varepsilon}_t,\end{aligned} \tag{32}$$

where $\boldsymbol{\varepsilon}_t \sim N_p(\mathbf{0},\boldsymbol{\Omega}\,)$, $t = 1,...,T$, $\boldsymbol{\delta}'= \boldsymbol{\psi}'\boldsymbol{\tau}_{\perp}\boldsymbol{\tau}_{\perp}'$ with $\boldsymbol{\psi}' = -(\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}$, $\boldsymbol{\zeta}'= \boldsymbol{\psi}'\boldsymbol{\tau}-\ \boldsymbol{\Omega}\boldsymbol{\alpha}_{\perp}(\boldsymbol{\alpha}_{\perp}'\boldsymbol{\Omega}\boldsymbol{\alpha}_{\perp})^{-1}(\boldsymbol{\alpha}_{\perp}'\boldsymbol{\Gamma}\boldsymbol{\beta},\boldsymbol{\xi})$ and $\boldsymbol{\xi}$ is defined in (5)

The relations $\boldsymbol{\beta}'\tilde{\mathbf{x}}_t + \tilde{\boldsymbol{\delta}}'\Delta\tilde{\mathbf{x}}_t$, with $\boldsymbol{\beta}' = [\boldsymbol{\beta}',\boldsymbol{\rho}_0,\boldsymbol{\rho}_{01}]$, $\tilde{\mathbf{x}}_t' = [\mathbf{x}_t', t, t_{91.1}]$, $\tilde{\boldsymbol{\delta}}' = [\boldsymbol{\delta}',\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_{01}]$, and $\Delta\tilde{\mathbf{x}}_t' = [\Delta\mathbf{x}_t', 1, D_s91.1]$, define $r$ stationary polynomially cointegrating relations, whereas the relations $\boldsymbol{\tau}'\Delta\tilde{\mathbf{x}}_t$ define $p - s_2$ stationary relations between the growth rates.

## 6.2 Linking $I(1)$ with $I(2)$

It is useful to see how the formulation (32) relates to the usual VAR formulation (3). Relying on results in Johansen (1997) the levels and difference components of the unrestricted VAR model (3) can be decomposed as:

$$
\begin{aligned}
\boldsymbol{\Gamma}\Delta\mathbf{x}_{t-1}+\boldsymbol{\Pi}\mathbf{x}_{t-1} &= (\boldsymbol{\Gamma}\overline{\boldsymbol{\beta}})\underbrace{\boldsymbol{\beta}'\Delta\mathbf{x}_{t-1}}_{I(0)} \\
&+(\boldsymbol{\alpha}\boldsymbol{\alpha}'\boldsymbol{\Gamma}\overline{\boldsymbol{\beta}}_{\perp 1}+\boldsymbol{\alpha}_{\perp 1})\underbrace{\boldsymbol{\beta}'_{\perp 1}\Delta\mathbf{x}_{t-1}}_{I(0)} \\
&+(\boldsymbol{\alpha}\overline{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}\overline{\boldsymbol{\beta}}_{\perp 2})\underbrace{\boldsymbol{\beta}'_{\perp 2}\Delta\mathbf{x}_{t-1}}_{I(1)} \\
&+\boldsymbol{\alpha}\underbrace{\boldsymbol{\beta}'\mathbf{x}_{t-1}}_{I(1)}
\end{aligned} \tag{33}
$$

where $\overline{\boldsymbol{\beta}}=\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}$ and $\overline{\boldsymbol{\alpha}}$ is similarly defined. The decomposition describes three types of linear relations between the growth rates, $\boldsymbol{\beta}'\Delta\mathbf{x}_{t-1}$, $\boldsymbol{\beta}'_{\perp 1}\boldsymbol{\Delta}x_{t-1}$and $\boldsymbol{\beta}'_{\perp 2}\Delta\mathbf{x}_{t-1}$, of which the first two define $I(0)$ relations and the third an $I(1)$ relation. The coefficients in soft brackets define the corresponding adjustment coefficients.

Since $\boldsymbol{\beta}'_{\perp 2}\Delta\mathbf{x}_{t-1}$ is $I(1)$, it needs to be combined with another $I(1)$ variable to become stationary. An obvious candidate for this is $\boldsymbol{\beta}'\mathbf{x}_{t-1}$. It is now easy to see how the parameterization in (3) relates to the one in (32):

$$
\boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{x}_{t-1}+(\overline{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}\overline{\boldsymbol{\beta}}_{\perp 2})\,\boldsymbol{\beta}'_{\perp 2}\Delta\mathbf{x}_{t-1})=\boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{x}_{t-1}+\boldsymbol{\delta}'\Delta\mathbf{x}_{t-1}). \tag{34}
$$

Finally, when $r>s_2$ the long-run matrix $\boldsymbol{\Pi}$ can be expressed as the sum of the two levels components measured by:

$$
\boldsymbol{\Pi}=\boldsymbol{\alpha}_0\boldsymbol{\beta}'_0+\boldsymbol{\alpha}_1\boldsymbol{\beta}'_1
$$

where $\boldsymbol{\beta}'_0\mathbf{x}_{t-1}$ defines $r-s_2$ directly stationary $CI(2,2)$ relations, whereas $\boldsymbol{\beta}'_1\mathbf{x}_{t-1}$ defines $s_2$ nonstationary $CI(2,1)$ cointegrating relations which needs to be combined with the differenced process to become stationary through polynomial cointegration.

Thus, the $I(2)$ model can distinguish between the $CI(2,1)$ relations between levels $\{\boldsymbol{\beta}'\mathbf{x}_t,\boldsymbol{\beta}'_{\perp 1}\mathbf{x}_t\}$, the $CI(1,1)$ relations between levels and differences $\{\boldsymbol{\beta}'\mathbf{x}_{t-1}+\boldsymbol{\delta}'\Delta\mathbf{x}_t\}$, and finally the $CI(1,1)$ relations between differences $\{\boldsymbol{\tau}'\Delta\mathbf{x}_t\}$. As a consequence, when discussing the economic interpretation of these components, the generic concept of a "long-run" equilibrium relation needs to be modified accordingly. Juselius (2006, Chapter 17) proposed the following interpretation:

- $\boldsymbol{\beta}'\mathbf{x}_t+\boldsymbol{\delta}'\Delta\mathbf{x}_t$ as $r$ *dynamic long-run equilibrium relations*, or alternatively when $r>s_2$
  - $\boldsymbol{\beta}'_0\mathbf{x}_t$ as $r-s_2$ *static long-run equilibrium relations*, and

– $\boldsymbol{\beta}_1'\mathbf{x}_t + \boldsymbol{\delta}_1\Delta\mathbf{x}_t$ as $s_2$ *dynamic long-run equilibrium relations,*

- $\boldsymbol{\tau}'\Delta\mathbf{x}_t$ as *medium-run equilibrium relations.*

# 7 Two hypothetical scenarios

To be able to structure and interpret the empirical VAR results, it is useful to formulate a scenario for what we would expect to find in the VAR model, provided the reality is in accordance with the assumption of the theoretical model. For example, the first scenario below is specified for the hypothesis: $\{ppp_t \sim I(0)$, prices are pushing and the nominal exchange rate is pulling$\}$ under the assumption that $\mathbf{x}_t$ is empirically $I(2)$.

We shall discuss the following two cases, (1) $r = 2$, which correponds to the theory consistent case, and (2) $r = 1$, which is what we find in the data. In both cases it will be assumed that long-run price homogeneity holds, i.e. $\boldsymbol{\beta}_{\perp 2}' = [c, c, 0]$.

**Case 1**: $\{r = 2, s_1 = 0, s_2 = 1\}$ is consistent with:

$$\begin{bmatrix} p_{1,t} \\ p_{2,t} \\ s_{12,t} \end{bmatrix} = \begin{bmatrix} c \\ c \\ 0 \end{bmatrix} \sum_{j=1}^{t}\sum_{i=1}^{j} u_{1,i} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \sum_{i=1}^{j} u_{1,i} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix} \tag{35}$$

It is easy to see that $(p_{1,t} - p_{2,t}) \sim I(1)$ and $(p_{1,t} - p_{2,t} - s_{12,t}) \sim I(0)$ if $(b_1 - b_2) = b_3$. When the nominal exchange rate is adjusting (and price shocks are pushing) one would have that $u_{1,t} = \boldsymbol{\alpha}_{\perp 1}'\boldsymbol{\varepsilon}_t$ with $\boldsymbol{\alpha}_{\perp 1}' = [a_1, a_2, 0]$. This scenario would imply two cointegrating relations, one of which is directly cointegrating, because $r - s_2 = 1$, and the other is polynomially cointegrating, because $s_2 = 1$. It is easy to show that the directly cointegrating relation is the *ppp* relation, i.e. $(p_{1,t} - p_{2,t} - s_{12,t}) \sim I(0)$. The polynomially cointegrated relation is more difficult to see and it is helpful to examine the system based on the nominal-to-real transformation (Kongsted, 2005)[8]:

$$\begin{bmatrix} p_{1,t} - p_{2,t} \\ \Delta p_{1,t} \\ s_{12,t} \end{bmatrix} = \begin{bmatrix} b_1 - b_2 \\ c \\ b_3 \end{bmatrix} \sum_{i=1}^{j} u_{1,i} + \begin{bmatrix} \tilde{\varepsilon}_{1,t} \\ \tilde{\varepsilon}_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

It is now straightforward to show that $\{p_{1,t} - p_{2,t} + \omega\Delta p_{1,t}\} \sim I(0)$, if $c = -(b_1 - b_2/\omega)$. Alternatively, if $c = -b_3/\omega$, then $\{s_{12,t} + \omega\Delta p_{1,t}\} \sim I(0)$. In both cases the polynomially cointegrating relation can be thought of as a

[8]From a statistical point of view, an equivalent transformation would be achieved by replacing $\Delta p_1$ with $\Delta p_2$.

dynamic equilibrium relation describing how the inflation rate adjusts when relative prices have been pushed apart, i.e. $\Delta p_{1,t} = -1/\omega\ (p_{1,t} - p_{2,t})$. It simply states the obvious that the inflation rates have to react in a non-homogeneous manner if relative prices move persistently apart.

**Case 2**: $\{r = 1, s_1 = 1, s_2 = 1\}$ is consistent with:

$$\begin{bmatrix} p_{1,t} \\ p_{2,t} \\ s_{12,t} \end{bmatrix} = \begin{bmatrix} c \\ c \\ 0 \end{bmatrix} \sum_{j=1}^{t} \sum_{i=1}^{j} u_{1,i} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{j} u_{1,i} \\ \sum_{i=1}^{j} u_{2,i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

*In this case there is not in general a directly cointegrating relation, as* $r - s_2 = 0$, *but one polynomially cointegrating relation, as* $s_2 = 1$. *Again, the properties can be more easily discussed in the nominal-to-real transformed system:*

$$\begin{bmatrix} p_{1,t} - p_{2,t} \\ \Delta p_{1,t} \\ s_{12,t} \end{bmatrix} = \begin{bmatrix} b_{11} - b_{21} & b_{12} - b_{22} \\ c & 0 \\ b_{31} & b_{32} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{j} u_{1,i} \\ \sum_{i=1}^{j} u_{2,i} \end{bmatrix} + \begin{bmatrix} \tilde{\varepsilon}_{1,t} \\ \tilde{\varepsilon}_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

*It is now easy to see that stationarity of* $ppp_t$ *can only be achieved in the very special case when* $b_{11} - b_{21} = b_{31}$ *and* $b_{12} - b_{22} = b_{32}$, *implying that* $\delta$ *in (32) takes the value zero. Generally,* empirical support for $ppp_t$ can only be achieved by polynomial cointegration, i.e. in the form of a dynamic long-run adjustment relation. For example, if $b_{12} - b_{22} = b_{32}$ and $c = -(b_{11} - b_{21} - b_{31})/\omega$, then $\{p_{1,t} - p_{2,t} - s_{12,t} + \omega \Delta p_{1,t}\} \sim I(0)$. The latter can be interpreted as evidence of the following dynamic adjustment relation: $\Delta p_{1,t} = -1/\omega \{p_{1,t} - p_{2,t} - s_{12,t}\}$. In this case, either inflation rates or the currency depreciation/appreciation rate have to move in an offsetting direction when *ppp* has persistently deviated from its benchmark values.

Thus, the outcome of testing rank indices in the $I(2)$ model has strong implications for whether support for a stationary relation can be found or not.

# 8 An $I(2)$ analysis of prices and exchange rates

## 8.1 Determining the two rank indices

The number of stationary multi-cointegrating relations, $r$, and the number of $I(1)$ trends, $s_1$, among the common stochastic trends, $p-r$, can be determined

Table 5: Determination of rank indices

| $r$ | $p$-$r$ | $s_2 = 3$ | $s_2 = 2$ | $s_2 = 1$ | | $s_2 = 0$ |
|---|---|---|---|---|---|---|
| 0 | 3 | 527.6 | 293.9 | 118.10 | | 80.06 |
| | | [110.9] | [89.3] | [71.9] | | [57.9] |
| 1 | 2 | | 96.88 | **32.25** | | **32.65** |
| | | | [64.4] | [48.5] | | [36.6] |
| 2 | 1 | | | **8.20** | | **6.72** |
| | | | | [28.7] | | [18.4] |
| The 4 largest characteristic roots, $r = 2$ | | | | | | |
| $s_1 = 0$ | $s_2 = 1$ | | 1.0 | 1.0 | 0.98 | 0.53 |
| The 4 largest characteristic roots, $r = 1$ | | | | | | |
| $s_1 = 2$ | $s_2 = 0$ | | 1.0 | 1.0 | 0.99 | 0.53 |
| $s_1 = 1$ | $s_2 = 1$ | | 1.0 | 1.0 | 1.0 | 0.53 |

Note: 95% quatiles in []

by the $ML$ procedure in Nielsen and Rahbek (2007), where the trace test is calculated for all possible combinations of $r$ and $s_1$ so that the joint hypothesis $(r, s_1)$ can be tested as explained below.

Table 5 reports the ML tests of the joint hypothesis of $(r, s_1)$ which corresponds to the two reduced rank hypotheses in (4) and (5). The test procedure starts with the most restricted model $(r = 0, s_1 = 0, s_2 = 3)$ in the upper left hand corner, continues to the end of the first row $(r = 0, s_1 = 3, s_2 = 0)$, and proceeds similarly row-wise from left to right until the first acceptance. Based on the tests, the first acceptance is at $(r = 1, s_1 = 1, s_2 = 1)$, which was also the preferred choice in Section 3. The last column of the table correponds to the $I(1)$ trace test. When the data are $I(2)$, determining the rank $r$ exclusively on this test can often lead to incorrect results.

Our model has a broken linear trend restricted to the polynomially cointegrated relation and a shift dummy restricted to the differences. Because of this, the standard asymptotic trace test distributions (for example, provided by CATS for RATS) are no longer correct. The critical values given in brackets below the test values have been kindly provided by Heino Nielsen using a simulation program described in Nielsen (2004): see also Kurita (2007). The inclusion of a broken linear trend in the cointegration relations shifts the distributions to the right, implying that the test will be undersized if one ignores the effect of the broken trend.

Table 5 also reports the characteristic roots in the VAR model for $r = 1$ and 2. For $\{r = 2, p - r = 1\}$ there is just one common stochastic trend, which has to be $I(2)$ if the data are $I(2)$. The choice of $\{r = 2, s_2 = 1\}$ will impose two unit root restrictions on the characteristic roots of the model. As already discussed in Section 4.2 and confirmed in Table 5, this leaves one large unrestricted root, 0.98, in the model. Such a root is not statistically distinguishable from a unit root and would give problems if left unrestricted in the empirical model. When $r = 1$, the choice $\{r = 1, s_1 = 1, s_2 = 1\}$ accounts for all three near unit roots in the model with 0.53 as the largest unrestricted root, whereas the choice of $\{r = 1, s_1 = 0, s_2 = 2\}$ corresponds to 4 unit roots in the model and basically forces 0.53 to be a unit root. Altogether, the results strongly suggest that $\{r = 1, s_1 = 1, s_2 = 1\}$ is the correct choice.

That $r = 1$ is an important result, as the two scenarios in Section 7 showed that a stationary $ppp_t$ is inherently associated with *one* stochastic trend having generated prices and nominal exchange rates. Thus, the finding of $p - r = 2$ suggests that there exists another source of permanent shocks that have contributed to the persistent behaviour in the data. A plausible explanation will be given in the concluding section.

## 8.2 The pulling forces

The scenarios above assume long-run price homogeneity. In Section 5, this hypothesis was tested on $\boldsymbol{\beta}'\mathbf{x}_t$ (see Johansen, 2006) and was accepted with high p-value. However, when $\mathbf{x}_t \sim I(2)$, long-run price homogeneity is defined on $\boldsymbol{\tau}'\mathbf{x}_t$, where $\boldsymbol{\tau}' = [\boldsymbol{\beta}, \boldsymbol{\beta}_{\perp 1}]$. Hence, long-run homogeneity on $\boldsymbol{\beta}$ is a necessary, but not sufficient condition. When tested, long-run price homogeneity of $\boldsymbol{\tau}'\mathbf{x}_t$ was strongly rejected based on $\chi^2(2) = 22.95[0.00]$ and $\boldsymbol{\beta}'_{\perp 1}\mathbf{x}_t$ cannot be considered homogeneous in prices. As a matter of fact, *the results in Table 6 demonstrate* that the coefficients to prices in $\boldsymbol{\beta}_{\perp 1}$ are proportional to (1, 1) rather than (1, -1). This, of course, is just another piece of evidence associated with the *ppp* puzzle.

Table 6 *also* reports the estimates of short-run adjustment dynamics towards the estimated long-run equilibrium relations. The $I(2)$ model is parameterized according to (32). We note that the $I(2)$ model allows the VAR variables to adjust to a medium-run equilibrium error, $\boldsymbol{\beta}'_{\perp 1}\boldsymbol{\Delta}\tilde{\mathbf{x}}_{t-1}$, to a change in the long-run 'static equilibrium' error, $\boldsymbol{\beta}'\Delta\tilde{\mathbf{x}}_{t-1}$, and to the long-run 'dynamic equilibrium' error, $\boldsymbol{\beta}'\tilde{\mathbf{x}}_{t-1} + \boldsymbol{\delta}\boldsymbol{\Delta}\tilde{\mathbf{x}}_{t-1}$. In this sense, the $I(2)$ model offers a much richer dynamic adjustment structure than the $I(1)$ model.

When discussing the adjustment dynamics with respect to the polynomially cointegrating relations, it is useful to interpret the adjustment coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ as two levels of equilibrium correction. Consider, for example, the following model for the variable $x_{i,t}$:

Table 6: The dynamics of the short-run dynamic adjustment

$$\underbrace{\begin{bmatrix} \Delta^2 p_{1t} \\ \Delta^2 p_{2,t} \\ \Delta^2 s_{12,t} \end{bmatrix}}_{I(0)} = \underbrace{\begin{bmatrix} \mathbf{-0.01} \\ {\scriptstyle [-4.98]} \\ \mathbf{-0.02} \\ {\scriptstyle [-8.66]} \\ \mathbf{-0.13} \\ {\scriptstyle [-3.41]} \end{bmatrix}}_{\boldsymbol{\alpha}} \underbrace{\begin{bmatrix} \boldsymbol{\beta}_1' \mathbf{x}_{t-1} + \boldsymbol{\delta}_1' \Delta \mathbf{x}_{t-1} \end{bmatrix}}_{I(0)}$$

$$+ \underbrace{\begin{bmatrix} \mathbf{-0.51} & \mathbf{-0.25} \\ {\scriptstyle [-11.97]} & {\scriptstyle [-11.15]} \\ \mathbf{0.29} & \mathbf{-0.14} \\ {\scriptstyle [7.25]} & {\scriptstyle [-6.51]} \\ 1.19 & 0.06 \\ {\scriptstyle [1.66]} & {\scriptstyle [0.15]} \end{bmatrix}}_{\boldsymbol{\zeta}} \underbrace{\begin{bmatrix} \boldsymbol{\beta}_1' \Delta \mathbf{x}_{t-1} \\ \boldsymbol{\beta}_{\perp 1,1}' \Delta \mathbf{x}_{t-1} \end{bmatrix}}_{I(0)} + \underbrace{\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}}_{I(0)}$$

where

$$\boldsymbol{\beta}_1' \mathbf{x}_t + \boldsymbol{\delta}_1' \Delta \mathbf{x}_t = \mathbf{1.0} p_{1,t} \underset{[-7.68]}{\mathbf{-0.85}} p_{2,t} + \underset{[15.08]}{\mathbf{0.19}} s_{12,t} \underset{[-5.99]}{\mathbf{-0.0025}} t_{91.1} + \underset{[8.34]}{\mathbf{0.0024}} t +$$
$$+ \mathbf{2.61}\ \Delta p_{1,t} + \mathbf{5.21}\ \Delta p_{2,t} + \mathbf{9.31}\ \Delta s_{12,t} - \mathbf{0.10}\ D_s 91.1$$

$$\boldsymbol{\beta}_{\perp 1}' \Delta \mathbf{x}_{t-1} = \mathbf{1.01}\ \Delta p_{1,t} + \mathbf{1.0}\ \Delta p_{2,t} - \mathbf{0.84}\ \Delta s_{12,t} + \mathbf{0.01}\ \Delta t_{91.1} - \mathbf{0.01}\ \Delta t$$

Note: t-ratios in []

$$\Delta^2 x_{i,t} = \cdots \sum_{j=1}^{r} \alpha_{ij} (\boldsymbol{\delta}_j' \Delta \mathbf{x}_{t-1} + \boldsymbol{\beta}_j' \mathbf{x}_{t-1}) + \cdots \tag{36}$$

If $\alpha_{ij}\delta_{ij} < 0$ for $j = 1, ..., r$, then the acceleration rates, $\Delta^2 x_{i,t}$, are equilibrium error correcting to the changes $\Delta x_{i,t}$, and if $\delta_{ij}\beta_{ij} > 0$ for $i = 1, .., p$, then the changes $\Delta x_{i,t}$, are equilibrium error correcting to the levels $x_{i,t}$. In the interpretation below we shall pay special attention to whether a variable is equilibrium error correcting or increasing as defined above, as this is an important feature of the data.

Based on the estimates in Table 6, it appears that the acceleration rates of prices and nominal exchange rates are all equilibrium error correcting to their respective growth rates. When it comes to the relationship between growth rates and levels of variables, there is just one polynomial cointegration relation to check for equilibrium correction, but the check has to be done for all three growth rates. To make the equilibrium correction property more visible, the relation $\boldsymbol{\delta}' \Delta \mathbf{x}_{t-1} + \boldsymbol{\beta}' \mathbf{x}_{t-1}$ has been formulated in three alternative, but equivalent, ways:

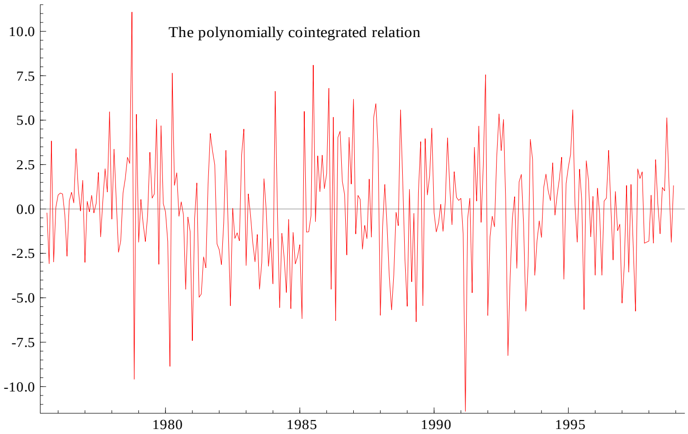


Figure 5: The graph of the polynomially cointegrated relation $\boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\delta}'\Delta\mathbf{x}_t$.

$$\Delta p_{1,t} = -\mathbf{0.38}(p_{1,t} - \underset{[-7.68]}{\mathbf{0.85}} p_{2,t} + \underset{[15.08]}{\mathbf{0.19}} s_{12,t} - \underset{[-5.99]}{\mathbf{0.0025}} t_{91.1} + \underset{[8.34]}{\mathbf{0.0025}} t) - \mathbf{2.0}\ \Delta p_{2,t} - \mathbf{3.5}\ \Delta s_{12,t}$$

$$\Delta p_{2,t} = \mathbf{0.16}(p_{2,t} - \underset{[-7.68]}{\mathbf{1.15}} p_{1,t} - \underset{[15.08]}{\mathbf{0.25}} s_{12,t} + \underset{[-5.99]}{\mathbf{0.003}} t_{91.1} - \underset{[8.34]}{\mathbf{0.003}} t) - \mathbf{0.50}\ \Delta p_{1,t} - \mathbf{1.8}\ \Delta s_{12,t}$$

$$\Delta s_{12,t} = -\mathbf{0.02}(s_{12,t} - \mathbf{4.5} p_{2,t} + \underset{[7.68]}{\mathbf{5.5}} p_{1,t} + \underset{[-5.99]}{\mathbf{0.013}} t_{91.1} - \underset{[8.34]}{\mathbf{0.013}} t) - \mathbf{0.28}\ \Delta p_{1,t} - \mathbf{0.56}\ \Delta p_{2,t}$$

It appears that the polynomially cointegrated relation is consistent with equilibrium correction behavior in the German inflation rate and the Dmk/$ depreciation/appreciation rate, whereas the US inflation rate is error increasing. The lack of equilibrium error correction in US prices, already commented on in Section 4.3, is an interesting empirical finding that is likely to be related to the *ppp* puzzle.

Ideally, one would like to interpret the above relations as dynamic adjustment of growth rates to a long-run static equilibrium relation, as described in the second scenario in Section 7. In the present case, this is not straightforward because the nominal exchange rate has the wrong sign in $\boldsymbol{\beta}'\mathbf{x}_t$. Therefore, the latter cannot be given an approximate interpretation of a long-run *ppp* relation. Whatever the case, Figure 5 illustrates that the polynomially cointegrated relation is strongly mean reverting.

Finally, the estimated adjustment coefficients, $\boldsymbol{\zeta} = [\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2]$, to the growth rate relations, $\boldsymbol{\beta}_1'\Delta\mathbf{x}_{t-1}$ and $\boldsymbol{\beta}_{\perp 1}'\Delta\mathbf{x}_{t-1}$, show that it is primarily the two prices that are adjusting. Both German and US prices are equilibrium adjusting to the first 'growth rates' relation, $\boldsymbol{\beta}_1'\Delta\mathbf{x}_{t-1} = 1.0\Delta p_{1t} - 0.85\Delta p_{2t} + 0.20\Delta s_{12,t}$, but German prices more quickly so. The second 'growth rates' relation, $\boldsymbol{\beta}_{\perp 1}'\Delta\mathbf{x}_{t-1} = 1.01\Delta p_{1,t} + 1.0\Delta p_{2,t} - 0.84\Delta s_{12,t}$ is more difficult to interpret. It essentially says that the change in the Dmk/$ rate has been proportional to the sum of German and US inflation rates, rather than to the inflation spread. As the coefficients of $\boldsymbol{\beta}_{\perp 1}$ are the opposite of price homogeneity, the results explain why long-run price homogeneity in $\boldsymbol{\tau}$ was so strongly rejected.

That inflation rates are moving in opposite directions is a puzzling and even implausible result. Therefore, it is useful to check whether this result still holds for the combined estimates, $\boldsymbol{\zeta}\boldsymbol{\tau}'\Delta\mathbf{x}_t$, calculated below:

| | $\Delta p_{1,t}$ | $\Delta p_{2,t}$ | $\Delta s_{12,t}$ |
|---|---|---|---|
| $\Delta^2 p_{1t}$ : | $-0.75$ | $0.18$ | $0.10$ |
| $\Delta^2 p_{2,t}$ : | $0.44$ | $-0.13$ | $0.04$ |
| $\Delta^2 s_{12,t}$ : | $1.25$ | $-1.00$ | $0.17$ |

Fortunately, the combined estimates are more plausible: German as well as US inflation rates are now equilibrium error correcting to each other. The US inflation rate is equilibrium error correcting to German price inflation with the correct sign but to the Dmk/$ rate with an 'incorrect' sign. However, the coefficient is very small and may not be significantly different from zero. Finally, the Dmk/$ rate is not equilibrium error correcting but even error increasing with the US-German inflation spread. Since the coefficients $\boldsymbol{\zeta}_{13}$ and $\boldsymbol{\zeta}_{23}$ were both insignificant this is, however, not necessarily an empirically strong result.

To summarize, the VAR analysis has detected four puzzling results:

1. Nominal exchange rates tend to move in the opposite direction to relative prices for extended periods of time.

2. The US inflation rate is not equilibrium error correcting to $\boldsymbol{\beta}'\mathbf{x}_t$.

3. Changes in the nominal exchange rate either do not seem to have been significantly responding to movements in relative inflation rates or, if they have, in an equilibrium increasing manner.

4. The US inflation rate does not seem to have been responding to this 'adverse' behavior of the change in the Dmk/$ rate.

## 8.3 The estimated driving forces

The scenario in Section 7 can now be directly assessed based on the estimates of the MA representation in Table 7. The results clearly show that the empirical reality has deviated quite substantially from the assumed theoretical scenario. For example, the estimated loadings to the $I(2)$ trend, $\boldsymbol{\beta}_{\perp 2}$, show that the price coefficients are not even close to being equal as assumed by the long-run homogeneity hypothesis. Given the previous rejection of long-run price homogeneity, this result should, of course, not come as a big surprise. However, what is more surprising is that the coefficient to the Dmk/$ rate is not even close to zero, suggesting that $s_{12,t}$ is empirically $I(2)$ rather than $I(1)$ as assumed in the scenario. Another surprising result is that, given the estimates of $\boldsymbol{\beta}_{\perp 2}$, the $I(2)$ trend does not seem to cancel in $ppp = p_1 - p_2 - s_{12}$. For this to be the case, the coefficients would need to be proportional to $\boldsymbol{\beta}'_{\perp 2} = [a, -a, 2a]$.

That the real exchange rate is empirically $I(2)$ would be hard to reconcile with standard theories. However, the theory of imperfect knowledge economics (Frydman and Goldberg, 2007) does in fact explain such a result. Frydman et al. (2008) demonstrate that, under highly plausible assumptions on agents' behavior, speculative transactions in the foreign exchange market are likely to generate pronounced persistence in nominal exchange rates that would be hard to distinguish from a near $I(2)$ process. Johansen et al. (2008) find strong evidence for this to be the case based on the same US-German data analyzed here, but extended with short- and long-term interest rates. They also find that the *ppp* transformed variable exhibits highly persistent behavior that can be considered either empirically near-$I(2)$ or $I(1)$, depending on whether emphasis is on size or power.

The estimated $\boldsymbol{\alpha}_{\perp 2}$ shows that it is shocks to relative prices (but with a larger weight on US prices) and to nominal exchange rates that seem to have generated the stochastic $I(2)$ trend. Contrary to the scenario, the coefficient to the nominal exchange rate is significant and the sign is opposite to the expected one. The estimated $\boldsymbol{\alpha}_{\perp 1,}$ describing the stochastic $I(1)$ trend, shows that a weighted average of inflationary shocks in Germany and the US have generated the medium run movements in prices and exchange rate.

These results seem to strengthen the previous conclusions: standard theories of price determination in the goods market cannot explain the *long swings in real exchange rate*s. The overriding impression it that it is the nominal exchange rate that is behaving oddly, suggesting that the long swings puzzle needs to be solved together with another international macro puzzle, the forward premium puzzle. This will be discussed in the concluding section.

Table 7: The common stochastic trends and their loadings

$$\begin{bmatrix} p_{1t} \\ p_{2,t} \\ s_{12,t} \end{bmatrix} = \begin{bmatrix} \mathbf{0.04} \\ \mathbf{0.09} \\ \mathbf{0.16} \end{bmatrix} \left[\boldsymbol{\alpha}'_{\perp 2,1} \sum \sum \hat{\boldsymbol{\varepsilon}}_s\right] +$$

$$+ \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}'_{\perp 2,1} \sum \hat{\boldsymbol{\varepsilon}}_i \\ \boldsymbol{\alpha}'_{\perp 1,1} \sum \hat{\boldsymbol{\varepsilon}}_i \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} \begin{bmatrix} t_{91.1} \\ t \end{bmatrix}$$

where

$$\boldsymbol{\alpha}'_{\perp 2,1} \hat{\boldsymbol{\varepsilon}}_t = \underset{[-4.03]}{-\mathbf{0.57}} \hat{\varepsilon}_{p_1,t} + \mathbf{1.0} \hat{\varepsilon}_{p_2,t} \underset{[-2.32]}{-\mathbf{0.09}} \hat{\varepsilon}_{s_{12},t}$$

$$\boldsymbol{\alpha}'_{\perp 1,1} \hat{\boldsymbol{\varepsilon}}_t = \underset{[6.79]}{\mathbf{0.25}} \hat{\varepsilon}_{p_1,t} + \underset{[2.52]}{\mathbf{0.14}} \hat{\varepsilon}_{p_2,t} \underset{[-1.82]}{-\mathbf{0.04}} \hat{\varepsilon}_{s_{12},t}$$

## 8.4 What did we gain from the $I(2)$ analysis?

Section 4 reported estimates and tests using the $I(1)$ model even though data were empirically $I(2)$. The question is whether the $I(2)$ analysis has changed some of the previous conclusions, or provided new insight that could not have been obtained from the $I(1)$ analysis.

To facilitate a comparison of the $I(1)$ and $I(2)$ models, it is useful first to subtract $\Delta \mathbf{x}_{t-1}$ from the both sides of the equation in (15) estimated in Section 4. The vector process would then be formulated in second differences $\Delta^2 \mathbf{x}_t$, and $\boldsymbol{\Gamma}_1$ would become $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_1 - \mathbf{I}$. In terms of likelihood, the two models differ only with respect to $\boldsymbol{\Gamma}$, which is unrestricted in the $I(1)$ model but subject to one nonlinear parameter restriction in the $I(2)$ model.

The estimates of the $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ coefficients are very similar in the two models, but their standard errors are smaller in the $I(2)$ model resulting in larger $t$ ratios. This is because in the $I(2)$ model the super-super consistency of $\boldsymbol{\beta}$ is adequately accounted for and because the $\boldsymbol{\beta}$ relation has been directly estimated as a polynomial cointegration relation. Also, the $\boldsymbol{\alpha}$ coefficients are not just measuring the adjustment to the levels relation, $\boldsymbol{\beta}' \mathbf{x}_{t-1}$, but to the levels and differences relation, $\boldsymbol{\beta}'_1 \mathbf{x}_{t-1} + \boldsymbol{\delta}'_1 \Delta \mathbf{x}_{t-1}$.

In the $I(1)$ model, the coefficient estimates of $\boldsymbol{\Gamma}_1$ are unrestricted, and there is not the same efficiency gain as in the $I(2)$ model, where the estimates are subject to the second reduced rank condition. In addition, the parametrization of the $I(1)$ model does not allow us to distinguish between $\boldsymbol{\beta}$ and $\boldsymbol{\tau} = (\boldsymbol{\beta}, \boldsymbol{\beta}_{\perp 1})$ and, therefore, not to decompose $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_1 - \mathbf{I}$ as in (33). Therefore, even though we may have realized that the $\boldsymbol{\beta}$ relation is not mean- reverting by itself and, thus, that it has to be combined with the differenced process $\boldsymbol{\delta}'\Delta\mathbf{x}_t$, we would not find the estimate of $\boldsymbol{\delta}$ without knowing the estimate of $\boldsymbol{\beta}_{\perp 1}$. Furthermore, the graphs of $\boldsymbol{\beta}_1'\mathbf{R}_{1,t}$ in Figure 3 and of $\boldsymbol{\beta}_1'\mathbf{x}_t + \boldsymbol{\delta}_1'\Delta\mathbf{x}_t$ in Figure 5 suggest that the latter relation is more precisely measured in terms of stationarity.

The hypothesis of long-run price homogeneity was adequately formulated as a test on $\boldsymbol{\tau}$ in the $I(2)$ model (and rejected), whereas in the $I(1)$ model it was formulated as a necessary, but not sufficient, test on $\boldsymbol{\beta}$ (and accepted). Thus, based on the $I(1)$ model, one might have been tempted to believe that long-run price homogeneity was acceptable even though it was strongly rejected. The rejection of homogeneity gave one of the clues as to why there *are puzzles in international economics*.

Finally, no useful results on the common driving trends could be obtained from the $I(1)$ model, whereas the MA analysis of the $I(2)$ model provided results on the $I(1)$ and $I(2)$ stochastic trends which suggested that we need to look closer at the determination of the nominal exchange rates.

To conclude, even though the $I(1)$ and $I(2)$ models are quite close in terms of likelihood, the $I(2)$ procedure is likely to insure against possible pitfalls in the statistical analysis when there is a double unit root in the data. Last, but not least, it also allows for a much richer structure and, therefore, more interesting interpretations of the information in the data.

# 9 Concluding discussion

The CVAR approach adopted in this chapter is based on general-to-specific modelling as a tool to uncover empirical regularities in the economy. Starting from a general unrestricted model representing the raw data and then testing down seems to be a useful way of extracting as much information as possible from the data without distorting them in a prespecified direction. In this vein, it is also important from the outset to untie any transformation of the variables, such as the *real exchange* transformation of prices and nominal exchange rates, assumed to hold rather than tested in the data. Such transformations, common in empirical economics, can often seriously distort signals in the data that, otherwise, might help to uncover important empirical regularities. This was also the case in this chapter, where the joint modelling of prices and exchange rates revealed empirical regularities in prices and the nominal exchange rate that were helpful in pinning down the underlying puzzling behavior in this period.

To effectively pull information from the data, this chapter argues that the vector process should be classified into directions of similar persistence, dubbed empirically $I(0)$, $I(1)$ or $I(2)$. By following this route, one can robustify inference and improve the interpretability of economic behavior in the short, medium and long run. However, the main advantage is the ability to associate persistent movements away from fundamental benchmark values in one variable/relation with similar persistent movements somewhere else in the economy. In a general equilibrium world one would expect a persistent imbalance in one sector to generate a persistent departure in another. Thus, by characterising the data according to the empirical order of integration, the CVAR approch offers a powerful tool to investigate the generating mechanisms underlying such puzzling behavior.

To distinguish between those empirical regularities which can be explained by the theory model and those which cannot, the chapter has demonstrated the importance of first translating the basic assumptions of the theory model into testable assumptions on the CVAR model. As an illustration, the chapter showed how to translate the assumption of a stationary PPP and long-run price homogeneity, together with the assumption that prices are pushing and the exchange rate is pulling, into testable hypotheses in the CVAR model. This theory consistent scenario showed, among others, that a stationary real exchange rate is inherently associated with *one* stochastic trend having generated prices and nominal exchange rates. The finding of two (rather than one) stochastic trends was particularly important, as it suggested the existence of an additional source of permanent shocks that have contributed to the persistent behaviour in the data. This additional shock seemed to be related to speculative behavior in the market for foreign exchange and pointed to the importance of addressing the PPP and the long swings puzzle jointly with another puzzle in international finance, the forward premium puzzle. Similar to the former, the forward premium puzzle also has to do with persistent movements in the data, now in the forward premium: $(R_{1,t} - R_{2,t} - E_t \Delta s_{12,t+m})$, where $R_{i,t}$ is an interest yield of maturity $m$.

Thus, the two puzzles have a common variable, the nominal exchange rate, suggesting that the puzzle is related to the joint determination of nominal exchange rates in the goods and the foreign exchange market. Based on a CVAR analysis of German and the US prices, exchange rates, and interest rates, Johansen et al. (2008) found that the *ppp* and the real interest rate spread were strongly cointegrating though individually I(1), or even near I(2). A theoretical justification for this strong feature in the data was provided by Frydman *et al.* (2008) who were able to show in a two-country monetary model with IKE that goods prices and exchange rates adjust to a long-run equilibrium relation being a combination of the *ppp* and the real interest rates spreads.

They also report additional findings that point to the importance of inflationary expectations measured by the term spread ($R_i^s - R_i^l$), which was found to be empirically $I(1)$. The latter finding, again, points to the importance of allowing for not just one, but at least two, stochastic trends in the term structure of interest rates (Giese, 2008), and thus to a reconsideration of the monetary policy interest rate channel.

This illustrates how the VAR approach can be used constructively. Starting with the basic information set, carefully structuring the information in the data, and adding more information if needed, might at an early stage suggest how to modify either the empirical or the economic model, or both.

The following passage from Hoover (2006) pinpoints the fundamental difference between an approach based on a priori theory and the general-to-specific approach to empirical economics:

> "The Walrasian approach is totalizing. Theory comes first. Empirical reality must be theoretically articulated before it can be empirically observed. There is a sense that the Walrasian attitude is that to know anything, one must know everything.
>
> ... There is a fundamental problem: How do we come to our a priori knowledge? Most macroeconomists expect empirical evidence to be relevant to our understanding of the world. But if that evidence only can be viewed through totalizing *a priori* theory, then it cannot be used to revise the theory.
>
> ... The Marshallian approach is archaeological. We have some clues that a systematic structure lies behind the complexities of economic reality. The problem is how to lay this structure bare. To dig down to find the foundations, modifying and adapting our theoretical understanding as new facts accumulate, becoming ever more confident in our grasp of the super structure, but never quite sure that we have reached the lowest level of the structure."

For example, the significant finding of two shocks rather than one and the rejection of long-run price homogeneity are two examples of important information in the data signalling the need to dig deeper in order to understand more. By taking this information in the data seriously, instead of just ignoring it, we have been able to uncover more structure and, thus, to improve our understanding, as demonstrated in Frydman et al (2008), Johansen et al. (2007) and Juselius (2008). Needless to say, the need to dig deeper does not stop here.

# 10 References

Bacchiocchi, E. and L. Fanelli (2005), Testing the purchasing power parity through $I(2)$ cointegration techniques. *Journal of Applied Econometrics* 20, 749–770.

Balassa, B. (1964), 'The purchasing parity doctrine: A reappraisal,' *Journal of Political Economy* 72, pp 584-596.

Cheung, Y-W and K.S. Lai (1993), 'Long-run purchasing power parity during the recent float', *Journal of International Economics*, Vol. 34, pp 181-92.

Colander, David (2006), ed. Post Walrasian macroeconomics: Beyond the dynamic stochastic general equilibrium model. Cambridge: Cambridge University Press.

Dennis, J., Johansen, S. and K. Juselius (2005), *CATS for RATS: Manual to Cointegration Analysis of Time Series*. Estima, Illinois.

Dornbusch, R. (1976), Expectations and exchange rate dynamics, *Journal of Political Economy*, December, 1161-1174.

Dornbusch, R. and J.A. Frankel (1988), The flexible exchange rate system: Experience and alternatives, in S. Borner, ed., *International Finance and Trade*, London: Macmillan, reprinted in J.A. Frankel ed., *On Exchange Rates*, Cambridge MA: The MIT Press, chapter 1,1995.

Froot, K. and Rogoff, K. (1995), Perspectives on $PPP$ and long-run real exchange rates', in *Handbook of International Economics*, Vol 3, (eds.), E. Grossman and K. Rogoff, Vol. 3 Amsterdam: North Holland.

Frydman, R. and M. Goldberg (2007), *Imperfect Knowledge Economics: Exchange Rates and Risk*. Princeton University Press, Princeton.

Frydman, R., Goldberg, M., K. Juselius, and S. Johansen (2008), Reexamining the purchasing power parity puzzle with imperfect knowledge economics, Preprint, University of Copenhagen.

Giese, J. (2008), Level, slope, curvature: Characterising the yield curve in a cointegrated VAR model. *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 2, 2008-28. http://www.economics-ejournal.org/economics/journalarticles/2008-28

Gonzalo, J. (1994), Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics*, 60, 203–233.

Hoover, K. (2006), The past as the future: The Marshallian approach to Post-Walrasian macro. In D. Colander (Ed.), *Post Walrasian Economics: Beyond the Dynamic Stochastic General Equilibrium Model*, Chapter 12. Cambridge University Press: Cambridge, Mass.

Hoover, K., Johansen, S. and Juselius, K. (2008), Allowing the data to speak freely: The macroeconometrics of the cointegrated vector autoregression. *Proceedings of the American Economic Association*, 2008.

Hoover, K. and Juselius, K. (2008), Re-examining the term structure of US constant maturity bond rates, Preprint, University of Copenhagen.

Johansen, S. (1992a), An $I(2)$ cointegration analysis of the purchasing power parity between Australia and the United States. In C. Hargreaves (Ed.), *Macroeconomic Modelling of the Long Run*, 229–248. Edward Elgar.

Johansen, S. (1992b), A representation of vector autoregressive processes integrated of order 2. *Econometric Theory* 8, 188–202.

Johansen, S. (1995a), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. 2.ed. 1996 Oxford University Press, Oxford.

Johansen, S. (1995b), A statistical analysis of cointegration for $I(2)$variables, *Econometric Theory* 11, 25-59.

Johansen, S. (1997), Likelihood analysis of the $I(2)$ model. *Scandinavian Journal of Statistics* 24, 433-462.

Johansen, S. (2005), The interpretation of cointegrating coefficients in the cointegrated vector autoregressive model. *Oxford Bulletin of Economics and Statistics* 67, 93-104.

Johansen, S. (2006), Confronting the economic model with the data," in Colander (2006), pp. 287-300.

Johansen, S. (2006), Statistical analysis of hypotheses on the cointegrating relations in the $I(2)$ model. *Journal of Econometrics* 132, 81–115.

Johansen, S. and Juselius, K. (1990), Maximum likelihood estimation and inference on cointegration. With application to the demand for money. Oxford Bulletin of Economics and Statistics, 52(2), 169-210.

Johansen, S. and Juselius, K. (1992), Testing structural hypotheses in a multivariate cointegration analysis of the $PPP$and the $UIP$ for UK, *Journal of Econometrics* 53, 211-244.

Johansen, S., Juselius, K., Frydman, R., and Goldberg, M. D. (2007) Testing hypotheses in an $I(2)$ model with applications to the persistent long swings in the Dmk/$ Rate," Forthcoming in Journal of Econometrics, 2007.

Juselius, K (1991), Long-run relations in a well defined statistical model for the data generating process: Cointegration analysis of the $PPP$ and $UIP$ relations between Denmark and Germany. In J. Gruber (ed.), *Econometric decision models: New methods of modeling and applications*, Springer Verlag, New York, NY.

Juselius, K. (1995), Do purchasing power parity and uncovered interest rate parity hold in the long run? An example of likelihood inference in a multivariate time-series model. *Journal of Econometrics*, 69, 211-240.

Juselius, K. (2006), *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press, Oxford.

Juselius, K. (2008), Persistent movements in PPP and UIP. An empirical or theoretical puzzle? Preprint, Economics Department, University of Copenhagen.

Juselius, K., and MacDonald, R. (2004), The International Parities Between USA and Japan. *Japan and the World Economy*, 16, 17–34

Juselius, K., and MacDonald, R. (2007), International parity relationships between Germany and the United States: A Joint Modelling Approach. In: A. Morales-Zumaquero (Ed.), *International Macroeconomics: Recent Development.* Nova Science Publishers.

Juselius, K. and Franchi, M. (2007), Taking a DSGE model to the data meaningfully, *Economics–The Open-Access, Open-Assessment E-Journal*, 4. http://www.economics-ejournal.org/economics/journalarticles/2007-4.

Juselius, K. and Johansen, S. (2006). Extracting information from the data: A European view on empirical Macro. In Colander (2006), pp. 301-334.

Kongsted, H. C. (2003), An $I(2)$ cointegration analysis of small country import price determiniation. *Econometrics Journal* 12, 163–178.

Kongsted, H. C. (2005), Testing the nominal-to-real transformation. *Journal of Econometrics* 124, 205–225.

Kongsted, H. C. and H. B. Nielsen (2004), Analyzing $I(2)$ systems by transformed vector autoregressions. *Oxford Bulletin of Economics and Statistics* 66, 379–397.

Kongsted, H. C., Rahbek, A., and C. Jørgensen (1999), Trend stationarity in the $I(2)$ cointegration model. *Journal of Econometrics* 90, 265–289.

Krugman, P. (1993), Exchange-Rate Instability. The MIT Press, Cambridge, Massachusetts.

Kugler, P. and C Lenz (1993), 'Multivariate cointegration analysis and the long-run validity of $PPP$', *Review of Economics and Statistics*, 75, 180-4.

Kurita, T. (2007), $I(2)$ cointegration analysis in the presence of deterministic shifts. Faculty of Economics, Fukuoka University.

Lothian, J. (1997), Multi-country evidence on the behaviour of purchasing power parity under the current float. *Journal of International Money and Finance.*

MacDonald, R. (1995), 'Long-run exchange rate modelling: a survey of the recent evidence', *International Monetary Fund Staff Papers*, 42.

MacDonald, R. and Marsh, I.W. (1999), *Exchange Rate Modelling*, Kluwer Academic Publishers.

MacDonald, R. and J Stein (1999), *Equilibrium Exchange Rates*, Kluwer Academic Publishers.

Nielsen, B. and Nielsen, H. B. (2006). The asymptotic distribution of the estimated charateristic roots in a second order autoregression. Preprint, Economics Department, University of Copenhagen.

Nielsen, H. B. (2004), Cointegration analysis in the presence of outliers. *The Econometrics Journal* 7, 249–271.

Nielsen, H. B. and A. Rahbek, (2007), The likelihood ratio test for cointegration ranks in the $I(2)$ Model. *Econometric Theory* 23, 615–637.

Nielsen, B. and A. Rahbek (2000) Similarity issues in cointegration Analysis. *Oxford Bulletin of Economics and Statistics* 62, 5–22.

Paruolo P. (2000), Asymptotic efficiency of the two stage estimator in $I(2)$ systems. *Econometric Theory* 16, 524–550.

Paruolo P. (2002), Asymptotic inference on the moving average impact matrix in cointegrated $I(2)$ VAR systems. *Econometric Theory* 18, 673–690.

Paruolo, P. and A. Rahbek (1999), Weak exogeneity in $I(2)$ VAR systems. *Journal of Econometrics* 93, 281–308.

Rogoff, K. (1996), The purchasing power parity puzzle. *Journal of Economic Literature* 34, 647-68.

Samuelson, P. (1964), 'Theoretical notes on trade problems' *Review of Economics and Statistics*, 46, pp 145-154.