Speed of Adjustment in Cointegrated Systems

Fanelli, Luca and Paruolo, Paolo

Department of Statistics, University of Bologna

June 2007

Speed of adjustment in cointegrated systems

Luca Fanelli[*], Paolo Paruolo[†]

June 29, 2007

Abstract

This paper considers the speed of adjustment to long-run equilibria, in the context of cointegrated Vector Autoregressive Processes (VAR). We discuss the definition of multivariate π-lives for any indicator of predictive ability, concentrating on cumulated interim multipliers which converge to impact factor for increasing forecasting horizon. Interim multipliers are related to autoregressive Granger-causality coefficients, structural or generalized cumulative impulse responses. We discuss the relation of the present definition of multivariate π-lives with existing definitions for univariate time series and for nonlinear multivariate stationary processes. For multivariate (possibly cointegrated) VAR systems, π-lives are functions of the dynamics of the system only, and do not depend on the history path on which the forecast is based. Hence one can discuss inference on π-lives as (discrete) functions of parameters in the VAR model. We discuss a likelihood-based approach, both for point estimation and for confidence regions. An illustrative application to adjustment to purchasing-power parity (PPP) is presented.

Keywords: π-life, speed of adjustment, impact factors, vector equilibrium correction, shock absorption.

J.E.L. Classification: C32, C52, F31.

[*]University of Bologna, Department of Statistical Sciences, via Belle Arti 41, I-40126 Bologna, Italy. E-mail: luca.fanelli@unibo.it.

[†]Corresponding author. Department of Economics, Via Monte Generoso 71, I-21100 Varese, Italy. Email: paolo.paruolo@uninsubria.it.

Contents

1 Introduction

Many economic relations and identification restriction schemes used in econometric analysis are formulated in terms of the long-run effect that a given variable (shock) exerts on another variable. A typical example is a neutrality restriction: under long-run monetary superneutrality, a permanent increase in the growth rate of the money stock should have no real effects – apart from real balances – in the long-run. On the contrary, endogenous growth models, such as Barro (1990), predict that government expenditure and taxation will have permanent (long-run) effects on economic growth.

Although economic theories are generally silent about the processes of adjustment to equilibria, in many circumstances they provide indications about the speed at which a given long-run effect must be reached. For this reason, measuring the speed of adjustment has attracted increasing attention among economists: purchasing power parity (PPP) is one of the leading examples. Half-lives are typical measures of speed of adjustment; they are usually defined in a univariate context, see e.g. Cheung and Lai (2000), Mark (2001), Kilian and Zha (2002) and Rossi (2005), inter alia.

The concept of adjustment is however most naturally stated in multivariate terms; this is the approach taken in Koop et al. (1996) who discuss impulse responses for nonlinear multivariate systems, and by Pesaran and Shin (1996, 1998) who propose persistence profiles and generalized impulse responses as indicators of speed of adjustment in cointegrated models. In these approaches the speed of convergence is inferred from impulse-response-type indicators, and no definition of multivariate half-life is given.

Recently, vanDijk et al. (2007) analyzed nonlinear system as Koop et al. (1996), and defined multivariate π-lives in this context. The present paper provides similar definitions for the case of cointegrated systems. We define a general indicator of cumulative effect of one variable on another, which contains also the cumulative impulse response (CIR) used e.g. in Andrews and Chen (1994) as a special case, and define multivariate π-lives for this indicator.

The present paper, which is in line with vanDijk et al. (2007), differs from it in several respects. First of all, due to the nature of cointegrated systems, we focus on the long-run response on the levels of variables (despite their nonstationarity), and we use the long-run effect as normalization for long-run π-lives. The explicit calculation of the long-run effect, which coincides with the definition of impact factor (IF) proposed in Omtzigt and Paruolo (2005), is possible because of the linearity-in-the-variables of the systems; this is not possible in general for nonlinear systems as the ones discussed in vanDijk et al. The present approach is discussed with special emphasis on I(1) systems, but it is directly applicable also to I(2) systems or to systems integrated of higher order.

Secondly, again due to linearity, here both the interim multipliers and the impact factor do not depend on the history path on which predictions are based. We are hence in the position to treat the π-lives as functions of the parameters only, and to address the problem of inference via likelihood methods as for any function of parameters. We find that the problem of constructing confidence intervals on the π-life is a nonstandard one, because the π-life we define is in general an integer. We address this problem by defining the set of π-life values that correspond to asymptotic confidence sets for the companion matrix, which is well-defined. The calculation of this confidence sets is non-trivial; we here propose a new algorithm suitable for this situation.

Thirdly, by focusing on the distinction between long-run and short-run properties of the system, we are able to distinguish different speeds of convergence, according to whether a given variable has significant long-run effects or not on the target variable. It particular when applying the definition of π-life to indicators of short-run speed, one finds cases discussed in vanDijk et al. (2007) when applied to linear systems. Moreover, the long-run $\frac{1}{2}$-life introduced here is shown to specialize to the univariate $\frac{1}{2}$-life in current use for univariate processes. Hence the present definition of π-life is a general one.

Our approach has direct connections to long-run Granger noncausality as defined in Dufour and Renault (1998) and Dufour et al. (2006). We show that long-run Granger noncausality implies a zero impact factor but not vice versa. We discuss the ensuing various possible cases, and observe that each one would be best described by a different choice of indicator, hence giving rise to different definitions of π-lives.

A special case of the indicator proposed here corresponds to cumulated structural or generalized impulse responses, see Koop et al. (1996). Thus the present approach covers all these impulse responses. Moreover, one may apply the present definition to persistence profiles or other measures based on the variance. However, also due to space constraints, we restrict attention here to impulse responses.

Our approach to the measure speed of adjustment can be applied to several fields of economic research. A typical example is consumption dynamics. Since most theories of aggregate consumption behavior suggest that consumption is smooth, and differ very little in terms of the predicted amount of consumption adjustment to shocks, Morley (2007) argues that a more powerful way to test e.g. the permanent income hypothesis (PIH) against habit formation and precautionary savings, is to determine whether consumption adjustment to equilibrium is fast (PIH holds) or slow (habit formation and precautionary savings hold).

PPP adjustment is another example. In the analysis of PPP adjustment, which is the area of investigation of the empirical illustration in Section 5, a relevant issue is whether nominal exchange rates or prices reverts faster to equilibrium, see Engel and Morley (2001), Cheung et al. (2004) and Crowder (2004). The PPP 'puzzle' is usually reported as the difficulty to reconcile the estimated half-life of PPP deviations, measured by the half-life of real exchange rates, with the observed price stickiness. If deviations from equilibrium have a monetary source, then the implied half-life should be no longer than one or two years, which is roughly the time it takes sticky goods prices and wages to adjust to monetary shocks; however, Rogoff's (1996) survey documents half-lives between three to five years for developed countries and the post-Bretton Woods period.[1] Sticky-price models, in the Dornbusch (1976) and Mussa (1982) tradition, stress the role of slowly adjusting prices in determining the reversion rate to equilibrium: given the differential speeds of adjustment characterizing asset markets and goods markets, the sluggishness of real exchange rates is directly tied to the speed of adjustment of nominal prices. The expected implication in this paradigm is that prices should adjust to PPP equilibrium not faster than nominal exchange rates.

The opposite view, recently supported by Engel and Morley (2001) and Cheung et al. (2004), maintains that the root of the PPP puzzle may lie in the possibly different speeds of convergence for nominal exchange rates and prices, and in particular that it is nominal exchange rates, not prices, that converge slowly toward PPP.

Another natural field of application is the one of policy effectiveness. When the policy maker may be able to set the value of some instrument variable (government expenditure, tax rate) with the aim of affecting a target variable, the impact factors defined in Omtzigt and Paruolo (2005) captures, *ceteris paribus*, the long-run impact of the intervention. Provided that the policy is effective, the speed at which the variable adjusts to its long-run level provide valuable information to the policy maker. One may envisage situations where the policy intervention that is accomplished more quickly is to be preferred over a similar intervention that would take longer to impact the variable of interest.

All these examples stress the importance of measuring whether a supposed long-run equilibrium effect is supported by the data, and the speed at which the convergence to equilibrium takes place. We argue that the concept of π-life provides a comprehensive tool to address the issue.

The rest of the paper is organized as follows. Section 2 presents the definition of π-life and the relations with the existing concepts of Granger-noncausality, impulse responses, shock absorption, univariate $\frac{1}{2}$-life. The proofs of this section are reported in Appendix A. Section 3 specializes these concepts to cointegrated systems of order 1 and defined IFs. Appendix B discusses connections of IF with Granger long–run noncausality in I(1) systems. Section 4 discusses likelihood-based inference on π-lives. Proofs of this section are reported in Appendix C, while additional formulae needed in the calculation of confidence sets are reported in Appendix D. Section 5 reports the illustration to PPP and Section 6 concludes.

2 Definitions

This section presents definitions. We choose a VAR framework with linear predictors and quadratic loss function. This choice allows us to concentrate attention on generalized CIR based on (possibly restricted) cointegrated VAR. Impulse responses are the object of a vast literature, see e.g. Lütkephol (1990), Sims and Zha (1999). The case of impulse responses in

[1] If PPP deviations were driven by real shocks alone, then it is would be hard to explain the high volatility of real exchange rates.

stationary nonlinear autoregressive processes is treated in Potter (2000) for the univariate case, and Van Dijk et al. (2007) for the multivariate one. In most of this section we present the problem along the lines of Omtzigt and Paruolo (2005), OP henceforth.

The rest of this section is organized as follows. Subsection 2.1 defines the forecasting problem and Subsection 2.2 defines a general multiplier for a given forecast horizon h; many measures in current use are linear combinations of this multiplier. In particular we characterize the relationships between this multiplier and the autoregressive causality-coefficients of Dufour and Renault (1998), as well as with structural and generalized impulse responses as defined in Koop et al. (1996). Subsection 2.3 defines (cumulative) interim multipliers and impact factors as defined in OP, and relates them to the present setup.

These indicators are used in Subsection 2.5 to define (multivariate) π-lives. Subsection 2.6 shows how the present definitions of π-life reduces to the usual definition for univariate processes, and discusses relation to shock-absorption measures.

2.1 The process

We assume that the observable variables at date t are collected in a $p \times 1$ vector X_t, which is generated by a VAR(k) process

$$\Pi(L)X_t = \mu^* D_t^* + \epsilon_t \tag{1}$$

where $\Pi(L) = I - \sum_{i=1}^{k} \Pi_i L^i$, ϵ_t is i.i.d. $N(0, \Omega)$, L is the lag operator, $\Delta := 1 - L$ is the difference operator, Ω is positive definite. The vector D_t^* represents a $d^* \times 1$ of vector deterministic component, like the constant. Unless otherwise stated, we assume $k \geq 2$ and we follow the notation used in Johansen (1996).

We assume that the roots of $|\Pi(z)| = 0$ satisfy $z = 1$ or $|z| > 1$. In particular, the case when there are no roots at $z = 1$ is called the I(0) case, see Johansen (1996). Recall also that X_t is called integrated of order j, $I(j)$, if $\Delta^j X_t$ is I(0) for $j = 1, 2, ...$ For the I(0), I(1) and I(2) cases (as well as in the general $I(j)$ case $j \in \mathbb{N}$), the system (1) can be represented in terms of a state vector $\widetilde{X}_t$ with a stable VAR(1) representation

$$\widetilde{X}_t = A\widetilde{X}_{t-1} + u_t \tag{2}$$

where $u_t := J\left(\mu^* D_t^* + \epsilon_t\right)$, $J := (I_p : 0)'$. Here $\widetilde{X}_t$ is $\widetilde{p} \times 1$ and A is $\widetilde{p} \times \widetilde{p}$ and stable, i.e. that all the eigenvalues of A are within the unit disk. The definition of the state vector $\widetilde{X}_t$ in the I(0) case is $\widetilde{X}_t := (X_t' : ... : X_{t-k+1}')'$. The I(1) case is described later in Section 3; for the I(2) case we refer to OP.

In the rest of this section we discuss definitions relative to the stable state-space representation (2); hence the given definitions apply generally to any systems (2).

In this section we use the stationary case with state vector $\widetilde{X}_t := (X_t' : ... : X_{t-k+1}')'$ for illustration purposes, and in order to connect the present concepts to the literature. Section 3 discusses application of the present concepts to the I(1) case.

2.2 Predictions and impulse responses

We consider the forecasting problem of $\widetilde{X}_{t+h}$ based on the information set $Z_t := \widetilde{X}^t_{-\infty} := (\widetilde{X}_{t-s}, s \geq 0)$, and consider a predictor $\widetilde{X}_{t+h|t} = g^\circ(h, Z_t) = g(h, \widetilde{X}_t)$, where g° and g represent appropriate functions. We concentrate for simplicity on the case of minimum mean-square error, linear predictor g, $\widetilde{X}_{t+h|t} = A^h \widetilde{X}_t$, which coincides with the conditional expectation $\mathrm{E}\left(\widetilde{X}_{t+h}|\widetilde{X}_t\right)$ for linear processes. In order to stress dependence of $\widetilde{X}_{t+h|t}$ on the value $\widetilde{x}$ of the conditioning variables, we write $\widetilde{X}_{t+h|t}(\widetilde{x})$ for $A^h\widetilde{x}$, the point predictor of $\widetilde{X}_{t+h}$ conditional on $\widetilde{X}_t = \widetilde{x}$.

We next consider changes in $\widetilde{x}$, from value $\widetilde{x}_1$ to $\widetilde{x}_2 := \widetilde{x}_1 + \widetilde{v}$. A measure of sensitivity of $\widetilde{X}_{t+h|t}(\widetilde{x})$ with respect to this change in $\widetilde{x}$ is given by

$$e(h, \widetilde{v}) := \widetilde{X}_{t+h|t}(\widetilde{x}_2) - \widetilde{X}_{t+h|t}(\widetilde{x}_1) = A^h \widetilde{v},$$

which is seen not to depend on the level of $\widetilde{x}_1$, but simply on the change $\widetilde{v}$ in $\widetilde{x}$, due to the linearity of the predictor $\widetilde{X}_{t+h|t}$ as a function of $\widetilde{x}$. This effect can be summarized by the $\widetilde{p} \times \widetilde{p}$ matrix coefficient

$$m\,(h) := \frac{\partial e(h, \widetilde{v})}{\partial \widetilde{v}'} = A^h.$$

This can be interpreted as a h-step ahead *multiplier* describing the effect of $\widetilde{v}$ onto $\widetilde{X}_{t+h|t}$.

Several indicators of forecast sensitivity are linear functions of $m(h)$. Specifically, Appendix A shows that linear functions of $m(h)$ include (i) structural impulse responses, (ii) generalized impulse response coefficients as defined in Koop et al. (1996), as well as (iii) autoregressive causality-coefficients defined in Dufour and Renault (1998), Dufour et al. (2006).

Take for instance structural IR. Let $\epsilon_t = B\eta_t$ where structural shocks η_t have expectation 0 and covariance I_p and B is square and nonsingular. Structural IR of $J'\widetilde{X}_t$ with respect to η_t are usually defined as the elements of $J'A^hJB$, which is seen to be a linear function of $m(h)$. As a further example, Appendix A shows that a subset of variables does not Granger-cause another subset of variables at horizon h if $m_{b,a}(h) := b'm(h)a = 0$ for appropriate choice of b and a. For later reference, the condition $m_{b,a}(h) = 0$ is called *Granger non-causality condition* at horizon h; if this condition holds for all h, we say it holds at all horizons. This concept is analyzed in more detail in Subsection 2.4 for the I(0) case; see Section 3 for the application of these concepts to I(1) systems.

Here we note that $m\,(h)$ is a generalization of the major sensitivity indicator of predictability. In the next Subsection we employ $m\,(h)$ to discuss long-run properties of forecasts, which have a direct interpretation for (co-)integrated systems.

2.3 Interim multipliers and impact factors

The h-step ahead multiplier $m(h)$ describes influence on forecasts h steps ahead. Given the stability of (2), however, one can calculate *cumulated* interim and total multipliers. In particular, consider the cumulated effect up to some horizon ℓ:

$$\mathrm{CE}(\ell, \widetilde{v}) := \sum_{h=1}^{\ell} e(h, \widetilde{v}) = \sum_{h=1}^{\ell} A^h \widetilde{v} = \Big((I - A^{\ell+1}) (I - A)^{-1} - I \Big) \widetilde{v}.$$

This effect can be summarized by the $\widetilde{p} \times \widetilde{p}$ matrix coefficient $F(\ell)$, called the *interim multiplier up to horizon* ℓ:

$$F(\ell) := \frac{\partial \mathrm{CE}(\ell, \widetilde{v})}{\partial \widetilde{v}'} = \sum_{h=1}^{\ell} A^h = (I - A^{\ell+1}) (I - A)^{-1} - I.$$

Because A is stable, as $\ell \to \infty$ the quantity $\mathrm{CE}(\ell, \widetilde{v})$ converges to a finite vector $(I - A)^{-1} \widetilde{v}$, called the total effect of $\widetilde{v}$, and the interim multiplier $F(\ell)$ converges to the limit

$$F(\ell) \underset{\ell \to \infty}{\to} F := \sum_{h=1}^{\infty} A^h = (I - A)^{-1} - I,$$

called the *total multiplier*, or *impact factor*, see OP, who note that $J' (F + I) J$ equals the CIR of X_{t+h} with respect to ϵ_t evaluated at ∞. The matrix coefficients $F(\ell)$ and F hence represent cumulated effects up to horizon ℓ or cumulated over all horizons. When some of the variables in $\widetilde{X}_t$ are for instance equal to ΔX_t – as will be the case for I(1) systems – the corresponding rows in $F(\ell)$ and F represent effects on the forecast of the levels $X_{t+\ell} - X_t$, see the discussion in OP and the following subsection.

Usually we are interested in the effect of a subset of variables x_t onto some other subset of variables y_t, where $x_t := \bar{a}' \widetilde{X}_t$ and $y_t := b' \widetilde{X}_t$ and a, b are known, user-defined, full-column-rank matrices. Here $\bar{a} := a(a'a)^{-1}$. It is simple to see that the cumulated effect of a change in x_t on the forecast up to ℓ periods ahead of y_t is given by $b' \mathrm{CE}(\ell, av)$, where the change $\widetilde{v}$ in $\widetilde{X}_t$ is given by $\widetilde{v} = av$. The size of the perturbation is represented by the Euclidean norm of $v = \bar{a}' \widetilde{v}$, $||v|| := (v'v)^{1/2}$. Note that the corresponding interim multiplier is $b' F(\ell) a$. In the following we use $||v||$ or $s\,||v||$ where s is a given scalar multiple, as possible denominator in order to normalize the interim multiplier $b' F(\ell) a$. In the rest of the paper a, b simply indicate selection vectors.

2.4 Long-run effects and Granger causality

In this subsection we discuss the relation between $F(\ell)$ and Granger-noncausality as discussed in Dufour and Renault (1998). It is observed that Granger-noncausality at all horizons implies an IF F equal to 0, but not vice versa. This suggests a classification of cases that is later used to discuss properties of different speeds of adjustment as measured by π-lives.

Consider a set of linear combinations b of the forecast variables $\widetilde{X}_{t+h}$ and some linear combination a of the conditioning variables $\widetilde{X}_t$; we let $F_{b,a}(\ell) := b' F(\ell) a$ and similarly $F_{b,a} := b' F a$ the corresponding linear combinations of multipliers. We say that $a' \widetilde{X}_t$ has a (cumulated) *long-run effect on* $b' \widetilde{X}_t$ if $F_{b,a} \neq 0$. We label this situation as 'Case 1'.

case	condition	description
1	$F_{b,a} \neq 0$	$a'\widetilde{X}_t$ has a long-run effect on $b'\widetilde{X}_t$
2	$F_{b,a} = 0$	$a'\widetilde{X}_t$ has no long-run effect on $b'\widetilde{X}_t$
2.1	$F_{b,a} = 0$ and $m_{b,a}(h) = 0$ for all $h = 1, ..., \infty$	$a'\widetilde{X}_t$ does not Granger-cause $b'\widetilde{X}_t$ at at all horizons *and hence* it has no long run effect on it
2.2	$F_{b,a} = 0$ and $m_{b,a}(h) \neq 0$ for some h	$a'\widetilde{X}_t$ Granger-causes $b'\widetilde{X}_t$ at some horizon *but* it has no long run effect on it

Table 1: Relations between presence of long-run effects and Granger-causality.

Note that one may have a long-run effect only when $a'\widetilde{X}_t$ *does* Granger cause $b'\widetilde{X}_t$ at *some* horizon $h \geq 0$.

Consider now the case $F_{b,a} = 0$, where a and b identify different blocks of variables. In this case there is no long-run effect, and we say that the effect is 'not permanent' or 'transitory'; we label this as 'Case 2'. The condition $F_{b,a} = 0$ is compatible with Granger non-causality of $a'\widetilde{X}_t$ on $b'\widetilde{X}_t$ (i.e. with the situation $m_{b,a}(h) = 0$ for all $h = 1, ..., \infty$), which we label 'Case 2.1'. It is also compatible with the situation where $a'\widetilde{X}_t$ Granger-causes $b'\widetilde{X}_t$, i.e. when $m_{b,a}(h) \neq 0$ for some h, but in such a way as to offset each other in the sum $F_{b,a} = 0$; we label this as 'Case 2.2'. These two situations are not distinguished in $F_{b,a} = 0$.

The preceding discussion shows that, while some variables may Granger-cause the variables of interest, this does not exclude the possibility of zero long-run effects. In this sense, the condition of zero long-run effect is less stringent than the one of absence of Granger-causality at all forecasting horizons. For ease of reference, we summarize Cases 1, 2.1 and 2.2 in Table 1.

This paper concentrates on Case 1; in this case, in fact, there is a long-run effect, and it makes sense to measure speed of adjustment with respect to this long-run effect. We define a version of π-life that is normalized on this long-run effect, called $N_\pi(F_{b,a}(\ell), F_{b,a})$ below.

Case 2 is also of (marginal) interest, as it characterizes all temporary effects. Given the absence of long-run effects, however, speed needs to be measured differently. In fact, it cannot be normalized on the long-run effect, given that this is equal to 0. To this purpose we entertain different definitions of π-life, which are normalized with respect to the size $s\,||v||$ of the perturbation; this is indicated as $N_\pi(F_{b,a}(\ell), s\,||v||)$ below.

2.5 The concept of π-life

In this section we discuss the definition of π-life in a multivariate context, using the interim and total multipliers $F(\ell)$ and F, as defined previously. We stress here that the concept of π-life as a measure of speed is *relative* to a given indicator. Hence we let $c(\ell)$ indicate a generic indicator, such as $m_{b,a}(\ell)$ or $F_{b,a}(\ell)$; Cases 1, 2, 2.1 and 2.2, originally defined for $F_{b,a}(\ell)$, are understood to be in terms of the generic indicator $c(\ell)$. When we need to refer to the complete sequence $c(\ell)$, $\ell = 1, 2, ...$ we indicate it as $\{c\} := \{c(\ell)\}_{\ell \in \mathbb{N}}$.

Consider first Case 1, where $F_{b,a} \neq 0$, i.e. $c(\infty) \neq 0$; one can normalize $c(\ell)$ relative to its long-run value $c(\infty)$. In other words, consider the ratio

$$\varphi_\ell := \frac{c(\ell)}{c(\infty)} - 1 \tag{3}$$

where note that φ_ℓ may also be negative. Because $c(\ell) \to c(\infty)$ as $\ell \to \infty$, one has $\varphi_\ell \to \varphi_\infty = 0$; note that φ_ℓ may oscillate wildly before converging to 0. Hence one can find the smallest forecast horizon $\ell - 1$ after which φ_ℓ stays permanently within an interval $\upsilon_\pi := [-\pi, \pi]$, with $\pi \in (0,1)$. The integer ℓ is then defined as the π*-life of the effect* $c(\ell)$, and it is indicated as $N_\pi(\{c\}, c(\infty))$ in the following.

Because φ_ℓ is a ratio, the fraction π in the approximation is relative to the final value $c(\infty)$. Hence the interpretation of the π-life is 'the forecast horizon after which $c(\ell)$ stays within $\pm$ a fraction π of its final value $c(\infty)$' and *not* the horizon at which a fraction π of the effect $c(\infty)$ has been accomplished. The leading choice of π is $\frac{1}{2}$, and one speaks of half-life, indicated as $N_{0.5}$. Note that $N_{0.5} \leq N_{0.25}$ or that $N_{\pi_1} \leq N_{\pi_2}$ for $\pi_1 > \pi_2$, because $[-\pi_2, \pi_2] \subset [-\pi_1, \pi_1]$.

One can express the definition of π-life through the use of the indicators, as in VanDijk et al. (2007). Consider in fact the indicator variable

$$\mathtt{I}_\pi(c(\ell), d) := 1\left(|c(\ell) - c(\infty)| \leq \pi\, |d|\right), \tag{4}$$

where $1(\cdot)$ is the indicator function. For Case 1, we are in particular interested in $\mathtt{I}_\pi(c(\ell), c(\infty))$, which takes value 1 if $-\pi \leq \varphi_\ell \leq \pi$ and 0 otherwise. We note that the formulation (4) of the event $-\pi \leq \varphi_\ell \leq \pi$ avoids ratios; this is preferable, because it implies that $\mathtt{I}_\pi$ is well defined also in Case 2, i.e. when $c(\infty) = 0$. Next define the composite indicator function

$$\mathtt{PI}_m^\pi(\{c\}, d) := \prod_{j=m}^{\infty} \mathtt{I}_\pi(c(j), d) \tag{5}$$

which signals with value 1 the event that all $\mathtt{I}_\pi(c(j), d)$ take on the value 1 from $j = m$ onwards. In other words, $\mathtt{PI}_m^\pi(\{c\}, d)$ equals one when $-\pi \leq \varphi_j \leq \pi$ for all $j \geq m$, i.e. iff φ_j has entered the $[-\pi, \pi]$ band definitively. The π-life $N_\pi(\{c\}, d)$ can then be defined as the integer

$$N_\pi(\{c\}, d) := \sum_{m=1}^{\infty} \left(1 - \mathtt{PI}_m^\pi(\{c\}, d)\right). \tag{6}$$

Note that $1 - \mathtt{PI}_m^\pi(\{c\}, d)$ contributes a 1 to $N_\pi(\{c\}, d)$ if φ_ℓ has not entered the $[-\pi, \pi]$ band definitively, and a 0 otherwise. In the following we often use the notation $N_\pi(c(\ell), d)$ in place of $N_\pi(\{c\}, d)$. In particular we are interested in $N_\pi(F_{b,a}(\ell), F_{b,a})$, which we call the 'long-run π-life'. This is designed for Case 1, even though it can be calculated also in Case 2.

Consider next Case 2, where $c(\infty) = 0$. The definition of $\mathtt{I}_\pi$ is also applicable in this case; more specifically $\mathtt{I}_\pi(c(\ell), c(\infty)) = 0$ if $c(\ell) \neq 0$ and $\mathtt{I}_\pi(c(\ell), c(\infty)) = 1$ if $c(\ell) = 0$. Next consider the Cases 2.1 and 2.2 in more detail. Take Case 2.1, where $c(\ell) = 0$ for all ℓ, which implies $c(\infty) = 0$. One has, $\mathtt{I}_\pi(c(\ell), c(\infty)) = 1$ for all ℓ and hence $\mathtt{PI}^\pi_m(\{c\}, c(\infty)) = 1$ for all m. This implies that $N_\pi(c(\ell), c(\infty)) = 0$ for all π. In particular for $c(\ell) = F_{b,a}(\ell)$, $d = c(\infty) = F_{b,a}$, there is Granger non-causality of $a'\widetilde{X}_t$ on $b'\widetilde{X}_t$ at all horizons in the present situation. One hence finds $N_\pi(F_{b,a}(\ell), F_{b,a}) = 0$, i.e. a π-life equal to 0.

In particular, this applies to cointegrated VAR(1) processes, when $a'\widetilde{X}_t = \beta' X_{t-1}$ and when the no-feedback condition $b'\alpha = 0$ holds. Take, as an example, the cointegrated VAR(1) process

$$\begin{cases} \Delta X_{1t} = -\frac{1}{2}(X_{1t-1} - X_{2t-1}) + \epsilon_{1t} \\ \Delta X_{2t} = \epsilon_{2t} \end{cases}$$

where $X_t := (X_{1t} : X_{2t})'$ is 2×1, $\beta = (1 : -1)'$. It can be easily recognized that for $a'\widetilde{X}_t = \beta' X_{t-1}$ and $b_2'\widetilde{X}_t = \Delta X_{2t}$, one has $F_{b_2,a} = 0$, $F_{b_2,a}(\ell) = 0$ all ℓ (Case 2.1), implying that $N_\pi(F_{b_2,a}(\ell), 0) = 0$ for all π, including $\pi = \frac{1}{2}$. On the other hand, for $a'\widetilde{X}_t = \beta' X_{t-1}$ and $b_1'\widetilde{X}_t = \Delta X_{1t}$ one has $F_{b_1,a} = -1$ (Case 1), and hence $N_{0.5}(F_{b_1,a}(\ell), F_{b_1,a}) = 2$. Hence $N_{0.5}(F_{b_1,a}(\ell), F_{b_1,a}) > N_{0.5}(F_{b_2,a}(\ell), F_{b_2,a})$, and one is lead to conclude that X_{2t} adjusts faster than X_{1t}, see e.g. Morley (2007).[2]

Consider now Case 2.2 with $c(\ell) \neq 0$ up to some horizon, $\ell_{\max}$ say, while $c(\infty) = 0$. In this case there is Granger-causality up to horizons $\ell_{\max}$, but no long-run effect; one has $\mathtt{I}_\pi(c(\ell), c(\infty)) = 0$, $\mathtt{PI}^\pi_m(\{c\}, d) = 0$ up to $\ell_{\max} - 1$, so that $N_\pi(c(\ell), c(\infty)) = \ell_{\max} - 1$. If $\ell_{\max} = \infty$ then $N_\pi(c(\ell), c(\infty)) = \infty$. Again one can specialize these results to $c(\ell) = F_{b,a}(\ell)$, $d = c(\infty) = F_{b,a}$, and note that one may expect very large π-lives in this case.

The value of N_π in Cases 2.1 and 2.2 is hence extreme: equal to 0 in Case 2.1 and possibly very large or equal to ∞ in Case 2.2. These extreme values are however not very meaningful, because indeed there is no long-run effect, $c(\infty) = 0$, and it makes little sense to 'normalize by 0'.

Hence for Case 2 one could consider the alternative solution of normalizing the cumulated interim multiplier $c(\ell)$ on the size of the perturbation $s\,||v||$, in the vein of VanDijk et al. (2007). This corresponds to the π-life $N_\pi(c(\ell), s\,||v||)$ and to substituting the ratio φ_ℓ with $\varphi^*_\ell := c(\ell) / (s\,||v||)$.

We call the π-lives $N_\pi(F_{b,a}(\ell), s\,||v||)$ or $N_\pi(m_{b,a}(\ell), s\,||v||)$ the 'short-run π-life'. Obviously $N_\pi(F_{b,a}(\ell), F_{ba})$ and $N_\pi(c(\ell), s\,||v||)$ are different measures, which are designed for cases 1 and 2 respectively. Of course they imply different π-lives.

2.6 Univariate processes and shock absorption

In this subsection we show that the definition of π-life given above reduces to the usual definition of half-life for univariate AR(1) processes and $\pi = \frac{1}{2}$. We next discuss differences and similarities of the present definition with shock absorption measures, as defined in vanDijk el al. (2007).

[2] One could argue that there is no adjustment of X_{2t} to $\beta' X_{t-1}$, and that measuring speed of adjustment is hence questionable here.

We first consider the univariate AR(1) case, X_t scalar with $A \neq 0$, $|A| < 1$. Obviously in this case only $a = 1$, $b = 1$ are the only possible choices, so we use $F(\ell)$ and F with no subscripts; we consider the half-life $N_{0.5}(F(\ell), F)$ as defined previously. Because $F = A/(1-A) \neq 0$, one can consider the ratio φ_ℓ with no loss of generality; one finds $F(\ell) = A\left(1 - A^\ell\right)/(1-A) = F\left(1 - A^\ell\right)$ and hence

$$\varphi_\ell = \frac{F(\ell)}{F} - 1 = 1 - A^\ell - 1 = -A^\ell,$$

where $|A| < 1$ by the stationary requirement. Hence $|\varphi_\ell| < \pi$ if and only if $\left|-A^\ell\right| < \pi$, where $\left|-A^\ell\right| = |A|^\ell$, and one finds $N_\pi(F(\ell), F) = \lceil \ln \pi / \ln |A| \rceil$; here $\lceil \cdot \rceil$ indicates the smallest greater integer function. We hence see that $N_{0.5}(F(\ell), F)$ delivers the usual notion of half-life, see e.g. Kilian and Zha (2002), Rossi (2005) and reference therein.

We next discuss differences of the present approach with π-lives as defined in vanDijk et al. (2007) in the context of shock absorption. We argue that these differences come naturally from the different contexts: here we discuss linear nonstationary systems, while vanDijk et al. (2007) are concerned with nonlinear stationary systems.

The first difference is that in nonlinear systems, π-lives N_π depend on the history path Z_{t-1} as well as on the values of the perturbation, here represented by $v = J'\widetilde{v}$. This is reflected e.g. in eq. (11) in vanDijk et al., where the π-life N_π is defined also as a function of the current shock to ϵ_t, which depends on $v = J'\widetilde{v}$, and of the information variables Z_{t-1}. Because of the present linear system approach, we find that N_π is independent of v and Z_{t-1}.

As a consequence Van Dijk et al. (2007) proceed by considering the distribution of N_π as a function of the random variables v and Z_{t-1} for fixed values of the autoregressive coefficients, and define appropriate summary measures of its distribution. In our context, N_π does not depend on v and Z_{t-1}, and we here treat $N_\pi(F_{b,a}(\ell), F_{b,a})$ as a function of A, the companion matrix. In practice, A needs to be estimated (see Section 4 below) and we address the inference problem of N_π as a (discrete) function of the parameters in A.[3]

A final third difference lies with the scaling of the forecast indicator. Van Dijk et al. (2007) use $d = ||v|| - c(\infty)$ in definition (4) above, while we prefer to scale $c(\ell) - c(\infty)$ by the terminal value itself $d = c(\infty)$. This choice is natural in the present context, because the speed of convergence is measured *relative to* the impact factor $c(\infty) = F_{b,a}$.

We next specialize the present definitions to the case of cointegrated I(1) systems.

3 Cointegrated systems

In this section we consider cointegrated I(1) systems in more detail, and apply the above definitions of π-life. It is well known, see Johansen (1996), that process (1) generates I(1) variables with no linear trend if the following conditions hold:

[3] We also allow a, b to possibly depend on other parameters like Ω, as it is the case for structural IR.

- I(1)_a: every root z of the characteristic polynomial of X_t satisfies $z = 1$ or $|z| > 1$.
- I(1)_b: $\Pi := -\Pi(1) = \alpha\beta'$, where α and β are $p \times r$ matrices of full rank $r < p$ and $\mu_1 = \alpha\beta_0'$ with β_0' a $r \times 1$ vector.
- $I(1)$_c: $\alpha_\perp' \Gamma \beta_\perp$ has full rank $p - r$, where $\Gamma := I - \sum_{i=1}^{k-1} \Gamma_i$.

We call these conditions the 'I(1) assumption'. Other specifications of the deterministic components can be considered as in Johansen (1996). We concentrate attention to this simple case, because it is the relevant one in the empirical illustration.

Under the I(1) assumption, the VAR can be written in (many equivalent) companion forms. Following OP, we let $\widetilde{X}_t := \left(\Delta X_t' : X_{t-1}'\beta : U_t'\right)'$ be the state vector, where $U_t := (\Delta X_{t-1}' : ... : \Delta X_{t-k+1}')'$ is of dimension $m \times 1$, and β is a basis of the cointegration space in Assumption I(1)_b. Furthermore, define $\Gamma_1^* := \alpha\beta' + \Gamma_1$, $\Phi_1 := \Gamma_2$, $\Phi_2 := (\Gamma_3 : ... : \Gamma_{k-1})$. The associated state space representation is

$$\widetilde{X}_t = A\widetilde{X}_{t-1} + u_t$$

with $u_t := J(\mu^* D_t^* + \epsilon_t)$, $J := (I_p : 0_{p \times m+r})'$, and

$$A := \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} := \begin{array}{c} \begin{array}{cccc} p & r & p & m-p \end{array} \\ \left(\begin{array}{cc|cc} \Gamma_1^* & \alpha & \Phi_1 & \Phi_2 \\ \beta' & I_r & & \\ \hline I_p & & & \\ & & I_{m-p} & \end{array}\right) \end{array} \begin{array}{c} \\ p \\ r \\ p \\ m-p \end{array} \tag{7}$$

where we have reported dimensions alongside blocks of the state vector and of the companion matrix. For brevity the A_{22} block in (7) is partitioned in blocks of p and $m - p$ rows times $m - p$ and p columns, unlike the other blocks. Zero entries are not reported unless when needed for clarity.[4]

We next recall that for the present choice of state vector, the first p rows of $F(\ell)$ and F can be associated with the level of $X_{t+\ell}$ and X_∞ respectively. Let in fact $\widetilde{x}_2 = \widetilde{x}_1 + \widetilde{v}$ and $x_i := J'\widetilde{x}_i$, $i = 1, 2$, $v := J'\widetilde{v}$; note that $J'\widetilde{X}_{t+h} = \Delta X_{t+h}$ and

$$\begin{aligned} J'\mathrm{CE}(\ell, \widetilde{v}) &= \sum_{h=1}^{\ell} J'\left(\widetilde{X}_{t+h|t}(\widetilde{x}_2) - \widetilde{X}_{t+h|t}(\widetilde{x}_1)\right) = \sum_{h=1}^{\ell} \Delta X_{t+h|t}(\widetilde{x}_2) - \sum_{h=1}^{\ell} \Delta X_{t+h|t}(\widetilde{x}_1) \\ &= \left(X_{t+\ell|t}(\widetilde{x}_2) - x_2\right) - \left(X_{t+\ell|t}(\widetilde{x}) - x_1\right) = \left(X_{t+\ell|t}(\widetilde{x}_1 + \widetilde{v}) - X_{t+\ell|t}(\widetilde{x}_1)\right) - v \end{aligned}$$

Hence one has

$$\begin{aligned} J'F(\ell) &= J'\frac{\partial \mathrm{CE}(\ell, \widetilde{v})}{\partial \widetilde{v}'} = \frac{\partial X_{t+\ell|t}(\widetilde{x}_1 + \widetilde{v})}{\partial \widetilde{v}} - J', \\ J'F &= \frac{\partial X_{\infty|t}(\widetilde{x}_1 + \widetilde{v})}{\partial \widetilde{v}'} - J', \end{aligned} \tag{8}$$

[4]Note that the companion form (7) is formulated for $k \geq 2$. This assumption is not restrictive from a representation point of view, because any VAR(1) can be written as VAR(2) with a zero second order matrix coefficient. OP discuss how the inference procedures should be modified to account for the case of a VAR(1) also in estimation and testing.

where by linearity we know that $\partial X_{t+\ell|t}\left(\widetilde{x}_1+\widetilde{v}\right)/\partial\widetilde{v}'$ does not depend on $\widetilde{x}_1$ In words, the first block of p rows of the interim multipliers $F(\ell)$ and of the impact factors F represent the variation induced onto the *levels* of the process by the changes $\widetilde{v}$ in $\widetilde{X}_t$. This observation was first made by Bedini and Mosconi (2000).

The form of F for I(1) systems has been derived in OP; this representation is relevant for hypothesis-testing on $F_{b,a}$. Under Assumption I(1)_a , the eigenvalues of A are less than 1 in modulus, and hence the companion matrix A in (7) is stable. OP show that the IF F is in this case given by

$$\begin{aligned} F+I &= \begin{pmatrix} B & B\begin{pmatrix}\psi\\ 0\end{pmatrix} \\ (i_{k-2}\otimes I:0)B & c_1+i_{k-2}\otimes C\psi \end{pmatrix} \\ &= \begin{pmatrix} C & (C\Gamma-I)\bar{\beta} & C\psi \\ \bar{\alpha}'(\Gamma C-I) & \bar{\alpha}'(\Gamma C\Gamma-\Gamma)\bar{\beta} & \bar{\alpha}'(\Gamma C-I)\psi \\ i_{k-2}\otimes C & i_{k-2}\otimes(C\Gamma-I)\bar{\beta} & c_1+i_{k-2}\otimes C\psi \end{pmatrix}, \end{aligned}$$

where

$$B:=\begin{pmatrix} C & (C\Gamma-I)\bar{\beta} \\ \bar{\alpha}'(\Gamma C-I) & \bar{\alpha}'(\Gamma C\Gamma-\Gamma)\bar{\beta} \end{pmatrix}$$

$c_1:=c_2\otimes I_p$, with c_2 a lower triangular matrix with ones on and below the main diagonal, $C=\beta_\perp\left(\alpha_\perp'\Gamma\beta_\perp\right)^{-1}\alpha_\perp'$, $\psi:=(\psi_2:...:\psi_{k-1})$, $\psi_i=\sum_{j=i}^{k-1}\Gamma_j$. This structure of F allows to formulate hypothesis like $F_{b,a}=0$ in terms of the parameters of the process for each choice of a and b. The relation between impact factors and long-run Granger causality is discussed in Appendix B.

4 Inference

In this section we describe how likelihood-based inference on π-lives can be obtained from corresponding likelihood-based inference on A, with special reference to the I(1) case. This is the relevant case in many applications, such as the one reported in Section 5.

The impact factors $F_{b,a}$ play a relevant role in the definition and normalization in the definition of $N_\pi\left(F_{b,a}(\ell),F_{b,a}\right)$. In particular the hypothesis

$$H_0: F_{b,a}=0, \tag{9}$$

can be tested before the estimation of $N_\pi\left(F_{b,a}(\ell),F_{b,a}\right)$. Tests of (9) can be performed as proposed in OP Section 6. Some of the hypotheses of the form (9) concern only the matrix C, and one can also use the testing approach described in Paruolo (1997), Section 7. Finally, sometimes (9) concern only the column space of C, and one can employ, inter alia, the LR tests proposed in Paruolo (2006). In the rest of this paper we assume that tests of (9) have been performed. If such tests do not yield a rejection, we advise to consider a short-run π-life of the form $N_\pi\left(F_{b,a}(\ell),||v||\right)$. If the test has yielded a rejection, one can consider the long-run π-life $N_\pi(F_{b,a}(\ell), F_{b,a})$ assuming (9) is false. The latter has been labelled Case 1 above, while the former Case 2. The rest of the paper focuses on long-run π-life $N_\pi(F_{b,a}(\ell), F_{b,a})$ under the assumption that $F_{b,a} \neq 0$.

4.1 Estimation

We consider the I(1) models defined in Johansen (1996) as the class of VAR processes (1) where $\Pi = \alpha\beta'$, with α and β matrices of dimension $p \times r$ and all other parameters are unrestricted, with Ω symmetric and positive definite. Among these models we concentrate on those which exclude trend-stationary behavior. In particular in the application we consider the model called H_3 in Johansen (1996), with $D_t^* = 1$ and μ_1 unrestricted, as well as model H_2 which is the submodel of H_3 where $\mu_1 = \alpha\rho_1$, with ρ_1 unrestricted.

Likelihood-based inference on the cointegration rank in these models is summarized in Johansen (1996) to which we refer for details. Once inference on the cointegration rank and on the specification of deterministic components is performed, these can be fixed in subsequent analysis.

Next one can test hypothesis on β, like $\beta = (1 : -1)'$. This is relevant for instance in applications to PPP such as the one reported in Section 5. If this test does not reject, one can impose $\beta = (1 : -1)'$. Otherwise the cointegrating vector β can be estimated unrestrictedly.

As it is well known, this estimator of β is superconsistent, so that β can be considered fixed in the definition of the companion matrix A; only $\widehat{\Gamma}_1^*$, $\widehat{\alpha}$, $\widehat{\Phi}_1$, $\widehat{\Phi}_2$ contribute to the first order asymptotic variance of $\widehat{A}$. In particular, let $\widehat{\beta}$ be the ML estimate of β described e.g. in Johansen (1996). Here the companion matrix $A = (G^{*\prime} : L')'$ in (7) is decomposed in the block of the first p rows, called G^*, and the block of the remaining $r + m$ rows, called L. The latter block L contains known values (zeros and ones) as well as β. It can be estimated by plugging-in $\widehat{\beta}$ for β, obtaining the estimator $\widehat{L}$. Next $\widehat{\beta}$ is substituted for β in the state vector $\widetilde{X}_{t-1} := \left(\Delta X_{t-1}' : X_{t-2}'\beta : U_{t-1}'\right)'$, obtaining the regressors $\widehat{X}_{1t}$. G^* is estimated from the regression

$$\Delta X_t = G^* \widehat{X}_{1t} + \text{constant} + \text{error}.$$

Finally $\widehat{A} := (\widehat{G}^{*\prime} : \widehat{L}')'$. OP Theorem 5 find that as $T \to \infty$

$$T^{1/2} H' \operatorname{vec}(\widehat{A}' - A') \xrightarrow{d} N(0, V), \qquad V := \Omega \otimes \Sigma^{-1}$$

where $\Sigma := \mathrm{E}\left(\left(\widetilde{X}_t - \mathrm{E}(\widetilde{X}_t)\right)\left(\widetilde{X}_t - \mathrm{E}(\widetilde{X}_t)\right)'\right)$. Here V is a positive definite matrix, $H := (J \otimes I_p) = (I_g : 0)'$ a known selection matrix with $g := p(p + r + m)$ columns, vec indicates the column stacking operator and $\xrightarrow{d}$ indicates convergence in distribution. V can be consistently estimated by the plug-in estimator $\widehat{V} := \widehat{\Omega} \otimes \widehat{M}_{11}^{-1}$ where $\widehat{\Omega} := \widehat{M}_{\epsilon\epsilon} := T^{-1}\sum_{t=1}^T \widehat{\epsilon}_t\widehat{\epsilon}_t'$, $\widehat{M}_{11} := T^{-1}\sum_{t=1}^T \widehat{X}_{1t}\widehat{X}_{1t}' - \left(T^{-1}\sum_{t=1}^T \widehat{X}_{1t}\right)\left(T^{-1}\sum_{t=1}^T \widehat{X}_{1t}\right)$.

We observe that $N_\pi\left(F_{b,a}(\ell), F_{b,a}\right)$ is a function of the companion matrix, for fixed a and b, which we express as $N_\pi\left(F_{b,a}(\ell), F_{b,a}\right) = h_\pi(A)$; a likelihood based estimator for the half-life is obtained as $\widehat{h}_\pi = h_\pi(\widehat{A})$ i.e. by substituting A with $\widehat{A}$ as an argument of the function $h_\pi(\cdot)$. $\widehat{h}_\pi$ is hence the likelihood-based, plug-in estimator of the π-life $N_\pi\left(F_{b,a}(\ell), F_{b,a}\right)$. The likelihood-based, plug-in estimator of the short-term π-life $N_\pi\left(F_{b,a}(\ell), ||v||\right)$ is defined similarly; we do not reflect the difference in the function h_π in the notation, but simply note that h_π represents a different discrete function for each π-life $N_\pi\left(F_{b,a}(\ell), F_{b,a}\right)$ or $N_\pi\left(F_{b,a}(\ell), ||v||\right)$.

4.2 Confidence intervals and tests

In this subsection we define confidence intervals for the π-lives $N_\pi\left(F_{b,a}(\ell), F_{b,a}\right)$ and $N_\pi\left(F_{b,a}(\ell), ||v||\right)$, where for the former we assume $F_{b,a} \neq 0$. We consider the ratio φ_ℓ defined in (3) as a function of A, $\varphi_\ell = \varphi_\ell(A)$. Proofs of propositions in this section are collected in Appendix B.

We first introduce some notation. Let $\mathcal{A}$ be a confidence set (an ellipsoid) for the companion matrix A, obtained using the asymptotic normality of $\widehat{A}$; specifically,

$$\begin{aligned}\mathcal{A} &:= \{A : T\operatorname{vec}(\widehat{A}' - A')' H \widehat{V}^{-1} H' \operatorname{vec}(\widehat{A}' - A') \leq \chi^2_{1-\eta}(g)\} \\ &= \left\{A : T\operatorname{tr}\left(\widehat{M}_{11}\left(\widehat{A} - A\right)' J \widehat{\Omega}^{-1} J' \left(\widehat{A} - A\right)\right) \leq \chi^2_{1-\eta}(g)\right\},\end{aligned}$$

where $\chi^2_{1-\eta}(g)$ is the $1-\eta$ quantile of a χ^2 distribution with g degrees of freedom. For large samples, $T \to \infty$, one has $\Pr(A \in \mathcal{A}) \to 1-\eta$. We assume that all values of $A \in \mathcal{A}$ are stable, a property that holds for large T if A is stable in the data generating process.[5]

Define also the set $\mathcal{H}_\pi := \{h_\pi(A), A \in \mathcal{A}\}$ as the set of all values of the π-life h_π obtained for any choice of $A \in \mathcal{A}$. In order to emphasize that the following proposition does not depend on convergence results, we state it for a confidence set $\mathcal{A}$ for which $\Pr(A \in \mathcal{A}) = 1-\eta$.

Proposition 1 *Let $\mathcal{A}$ be a confidence set for A, i.e.* $\Pr(A \in \mathcal{A}) = 1-\eta$. *Let the set $\mathcal{H}_\pi := \{h_\pi(A), A \in \mathcal{A}\}$ be the corresponding set of values h, where h is any measurable function of A, possibly discrete. Then*

$$\Pr(N_\pi \in \mathcal{H}_\pi) \geq \Pr(A \in \mathcal{A}) = 1-\eta,$$

i.e. $\mathcal{H}_\pi$ is a confidence set for h with coverage probability at least equal to $1-\eta$.

The above proposition defines $\mathcal{H}_\pi$ as a confidence set for the π-lives $N_\pi\left(F_{b,a}(\ell), F_{b,a}\right)$ and $N_\pi\left(F_{b,a}(\ell), ||v||\right)$. The min and max values in $\mathcal{H}_\pi$, called $h^\pi_{\min}$, $h^\pi_{\max}$ provide bounds for

[5] For finite sample, this may not be the case, i.e. some of matrices $A \in \mathcal{A}$ may have eigenvalues on or outside the unit disk. In this case $F_{b,a}(\ell)$ may be undefined (if some of the roots of A are equal to 1), and if it is, $F_{b,a}(\ell)$ will generally fail to converge, and hence $F_{b,a}$ does not exist and/or it is $\pm\infty$. In this case, in the empirical illustration we conventionally assign value ∞ to $N_\pi(F_{b,a}(\ell), F_{b,a})$.

the π-life N_π, with assigned coverage probability $\geq 1-\eta$. Note that, unlike confidence intervals for impulse responses $\mathrm{IR}(\ell)$ calculated by the δ-method, see e.g. Lütkepohl (1990), which hold pointwise for fixed ℓ, the confidence set $h^\pi_{\min} \leq N_\pi \leq h^\pi_{\max}$ delivers a coverage probability of $1-\eta$.

In order to use $\mathcal{H}_\pi$ in practice, one is left with the problem of how to calculate $\mathcal{H}_\pi$. The problem is that $\mathcal{A}$ is uncountable, and a direct grid search may be unfeasible in many dimensions. Note that $h_\pi(A)$ is a discrete function, and hence it is not differentiable as a function of A; hence one cannot apply Newton-like methods directly to it.

However, one can find extreme values of $\varphi_\ell = \varphi_\ell(A)$ in (3) as a function of $A \in \mathcal{A}$. Specifically, fix $\ell \in \mathbb{N}$; the optimization problems

$$\varphi_{\ell,\min} := \min_{A\in\mathcal{A}} \varphi_\ell(A), \qquad \varphi_{\ell,\max} := \max_{A\in\mathcal{A}} \varphi_\ell(A) \tag{10}$$

are well defined over the compact set $\mathcal{A}$ and have a global minimum and maximum. Let $A_{\ell,\min} := \arg\min_{A\in\mathcal{A}} \varphi_\ell(A)$, $A_{\ell,\max} := \arg\max_{A\in\mathcal{A}} \varphi_\ell(A)$ denote the values of A that optimize (10). Note that the subset of values which includes $A_{\ell,\min}$, $A_{\ell,\max}$, $\ell \in \mathbb{N}$, is a countable subset of $\mathcal{A}$. We exploit this subset to deduce information on the location of $h^\pi_{\min}$ and $h^\pi_{\max}$ in the following way.

Consider the pair $\varphi_{\ell,\min}, \varphi_{\ell,\max}$ and the associated interval $\kappa_\ell := [\varphi_{\ell,\min}, \varphi_{\ell,\max}]$, which we compare with the interval $\upsilon_\pi := [-\pi, \pi]$. For large values of ℓ, we know that both $\varphi_{\ell,\min}$ and $\varphi_{\ell,\max}$ converge to 0, because of the assumed stability of all $A \in \mathcal{A}$. Hence κ_ℓ becomes a subset of υ_π for large ℓ. We can picture the relation between κ_ℓ and υ_π drawing a graph of κ_ℓ and υ_π against ℓ. υ_π describes a horizontal band around the ℓ axis, plus or minus π. κ_ℓ instead describes a sequence of intervals, whose length and whose endpoints all converge to 0 for $\ell \to \infty$.

In order for the κ_ℓ intervals to become subsets of υ_π for large ℓ, they need to have nonempty intersection with it. One can hence compute the first lead time ℓ_1 at which $\kappa_j \cap \upsilon_\pi \neq \emptyset$ in the following way

$$\ell_1 := \sum_{m=1}^{\infty} \prod_{j=1}^{m} 1\left(\kappa_j \cap \upsilon_\pi \neq \emptyset\right).$$

Similarly one can compute the smallest value of ℓ at which $\kappa_{\ell+j} \subseteq \upsilon_\pi$ for all $j = 0, 1, 2, ...$ as follows

$$\ell_2 := \sum_{m=1}^{\infty} \left(1 - \prod_{j=m}^{\infty} 1\left(\kappa_j \subseteq \upsilon_\pi\right)\right).$$

The following proposition shows that ℓ_1 and ℓ_2 convey valuable information on $h^\pi_{\min}$ and $h^\pi_{\max}$.

Proposition 2 *One has $h^\pi_{\min} \geq \ell_1$ and $h^\pi_{\max} = \ell_2$. Hence ℓ_1, ℓ_2 define conservative bounds on the π-life N_π, i.e. as $T \to \infty$*

$$\Pr\left(\ell_1 \leq N_\pi \leq \ell_2\right) \to \gamma \geq 1-\eta.$$

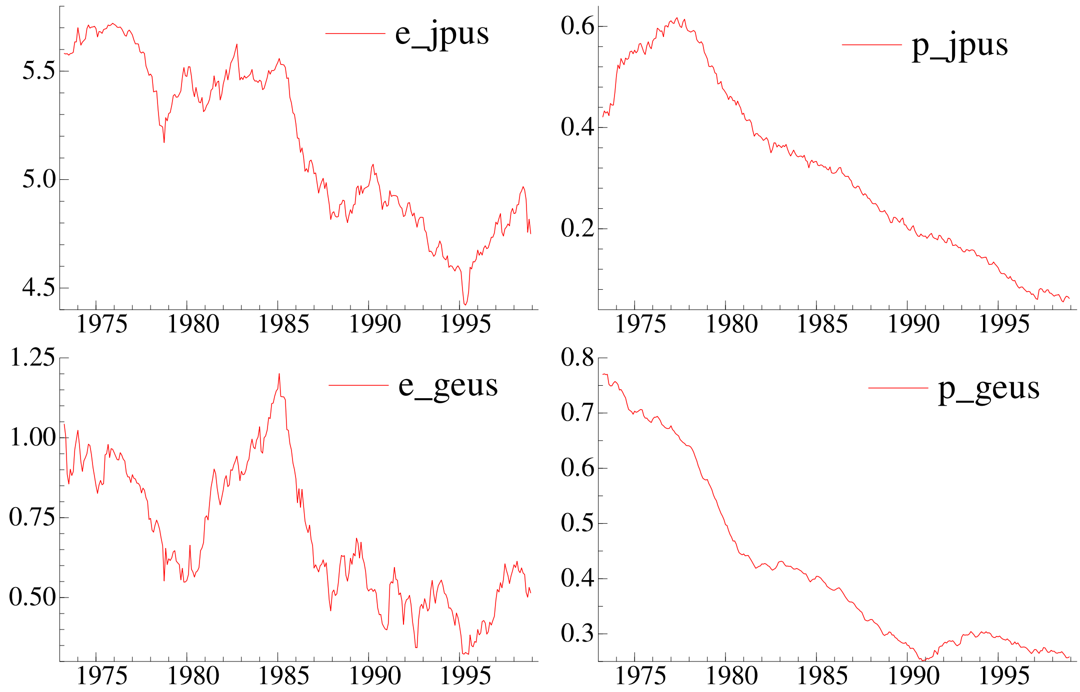

Figure 1: Levels of e_t and p_t for JP and GE.

This proposition allows to find $h^{\pi}_{\max}$ and the approximate location of $h^{\pi}_{\min}$ using the extreme values of the optimization problems (10). In order to show that $h^{\pi}_{\min} = \ell_1$ and not $> \ell_1$, it is enough to find a value of $A \in \mathcal{A}$ for which $h_{\pi}(A) = \ell_1$; trial values for A are provided by $A_{\ell_1,1}$, or the values of A visited by the Newton-like algorithm in the search for $A_{\ell_1,1}$. One may decide to simply compute ℓ_1, ℓ_2 as confidence bounds for N_{π}, or to investigate further if one can find an $A \in \mathcal{A}$ for which $h_{\pi}(A) = \ell_1$, so that to establish if $h^{\pi}_{\min} = \ell_1$.

Both solutions are based on the optimization problems (10). In practice it is sufficient to solve them for $\ell = 1, ..., \ell_{\max}$, for a suitably large $\ell_{\max}$. In Appendix C we report relevant derivatives of $\varphi_{\ell}(A)$ that facilitate application of Newton-like methods to (10).

5 An illustration

In this section we illustrate empirically how the π-life defined in equation (6) and the inference methods discussed in Section 2 can be applied to measure the speed of adjustment of nominal exchange rates and prices to PPP, for suitable choices of $b'\widetilde{X}_t$ and $a'\widetilde{X}_t$ (and, if required, for a given structuralization of VEC shocks, see Appendix A). We focus on the two most heavily traded exchange rates pairs during the period 1973-1998, namely dollar-deutschmark and dollar-yen. The monthly exchange rates and relative prices for the period 1973-1998 for these two country pairs are presented in Figure 1 and 2, along with their first

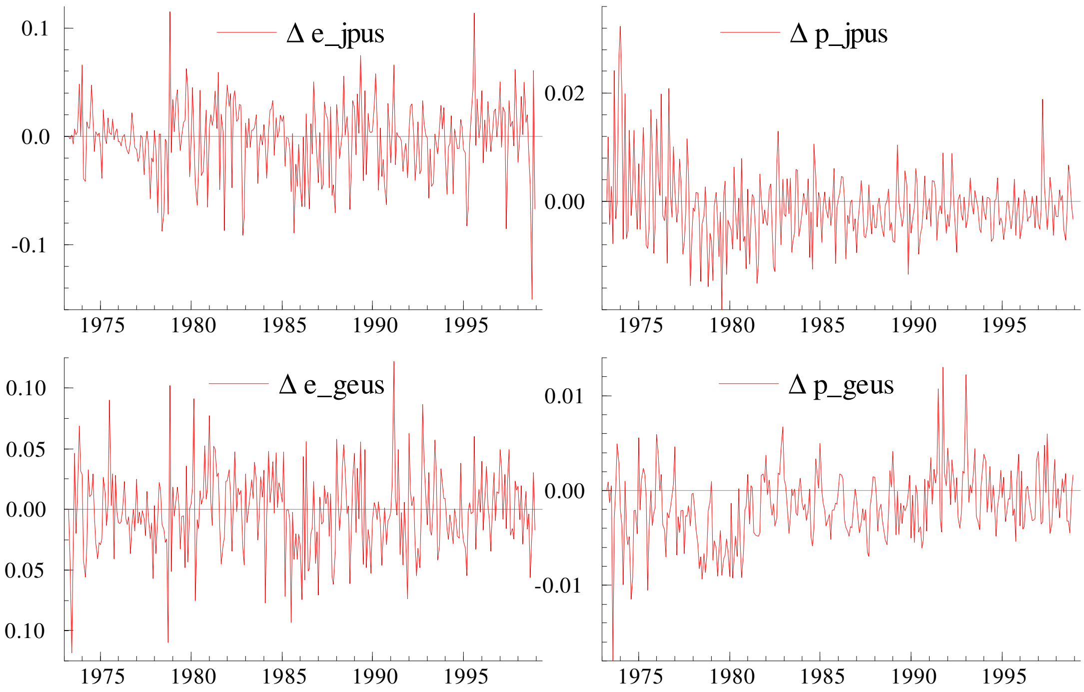

Figure 2: Δe_t, Δp_t for JP and GE.

differences.[6] Calculations were performed in Gauss and Ox.

We consider an I(1) bivariate VAR(2) system for $X_t = (e_t : p_t)'$, where e_t is the log of the nominal exchange rate (domestic vs foreign currency), and $p_t := p_t^d - p_t^{\mathrm{US}}$ is the log of relative prices (domestic vs foreign), with p_t^d the log of the CPI index, $d =$GE, JP, US. When also the real exchange rate $q_t = e_t - p_t$ is I(0), as predicted by PPP, one can define the state vector $\widetilde{X}_t = (\Delta e_t : \Delta p_t : q_{t-1})'$. In this case, given e.g. $b' = (1 : 0 : 0)$ and $a' = (0 : 0 : 1)$, the quantity $F_{b,a} = b' \sum_{h=0}^{\infty} A^h a = b'(F + I)a$ captures the long-run response of the nominal exchange rate e_t (and *not* the depreciation rate Δe_t) to variations in the real exchange rate q_t. Accordingly, in this set-up the π-life of $F_{b,a}$ represents a measure of speed for the adjustment of the nominal exchange rate to PPP deviations.

For easy of reference throughout this section we shall use the notation $F_{b_1,a} := F_{e,q}$ and $F_{b_2,a} := F_{p,q}$, where $b_1' = (1 : 0 : 0)$, $b_2' = (0 : 1 : 0)$ and a is defined as above. Other IFs of interest in this application are $F_{a,b_1} := F_{q,\Delta e}$ and $F_{a,b_2} := F_{q,\Delta p}$, which capture the 'permanent' (or 'transitory', if equal to zero) response of the real exchange rate to variations in the depreciation rate and in the inflation differential, respectively. Finally, $F_{a,a} := F_{q,q}$ can be regarded as the long run response of the real exchange rate to composite variations in nominal exchange rates and prices. Note that $N_{0.5}(F_{q,q}(\ell), F_{q,q})$ is usually interpreted in

[6] Nominal exchange rates are expressed as national currency units per 1 U.S. dollar; prices, which are measured in terms of CPI indices, are seasonally unadjusted and have base year 2000. Data are taken from the International Monetary Fund's IFS on-line database, and cover the period 1973.04–1998.12, prior to the introduction of the Euro. See http://ifs.apdi.net/imf/logon.aspx.

the literature on PPP as the half-life of real exchange rates.

Using the half-life as criterion, and provided that both $F_{e,q}$ and $F_{p,q}$ are different from zero (Case 1), the comparison of the π-lives $N_{0.5}(F_{e,q}(\ell), F_{e,q})$ and $N_{0.5}(F_{p,q}(\ell), F_{p,q})$ allows to establish whether nominal exchange rates or prices revert faster to equilibrium in response to PPP deviations. Conversely, the comparison of $N_{0.5}(F_{q,\Delta e}(\ell), F_{q,\Delta e})$ with $N_{0.5}(F_{q,\Delta p}(\ell), F_{q,\Delta p})$ reveals whether PPP deviations (real exchange rates) adjust faster in response to exchange rate depreciations or inflation differentials, respectively.[7]

5.1 Cointegration analysis

As in Cheung et al. (2004), we use cointegrated VECs of the form $X_t := (e_t : p_t)'$. Although in principle VAR shocks can be opportunely orthogonalized (see Appendix A), in this case the analysis does not take any theoretical stand on the process of adjustment driving exchange rates and prices. More precisely, we do not impose any specific structural restrictions other than the long-run PPP condition, as in Cheung et al. (2004).

The simple graphical inspection of Figure 1 suggests that nominal exchange rates peak around 1985. Although there is not a general consensus among economists, a shift in the policy regime towards a more active stance in managing external imbalances through policy coordination might have occurred in the aftermath of the Plaza Agreement of September 1985, see e.g. Klein et al. (1991). For this reason, before investigating the speed of PPP reversion of nominal exchange rates and prices, we apply Hansen's (2003) test for structural changes to establish whether the Plaza Agreement, other than representing a watershed in the active management of exchange rates among industrialized countries, also changed the structure of PPP adjustment.

	lnLik model with break in $\alpha\beta'$	lnLik model no break	LR	df	p-value
GE/US	1934.94	1928.75	12.38	4	0.015
JP/US	1717.92	1709.05	17.74	4	0.001

Table 2: Hansen's (2003) test for structural changes in the cointegrated VECs. Full Sample: 1973.04-1998.12, location of break: 1985.09, model H_2, VAR(2), $r = 1$.

Hansen's (2003) LR tests, reported in Table 2, compare the likelihood of the cointegrated VEC (e_t and p_t are cointegrated) where the matrices α and β do not change before and after the break date (restricted model), with the likelihood of the VEC where the matrices α and β (and the covariance matrix Ω) take different values before and after 1985.09 (unrestricted model). Each estimated VEC includes $k = 2$ lags and a constant restricted

[7] Note that differently from what recently claimed in the literature, where it is reported that VECs imply the same speed of adjustment to equilibrium for the variables (Morley, 2007), there is no reason for $N_{0.5}(F_{e,q}(\ell), F_{e,q})$ and $N_{0.5}(F_{p,q}(\ell), F_{p,q})$ (or $N_{0.5}(F_{q,\Delta e}(\ell), F_{q,\Delta e})$ and $N_{0.5}(F_{q,\Delta p}(\ell), F_{q,\Delta p})$) being equal.

	Model	$r=0$	$r \leq 1$	$\lambda_i(\widehat{A})$, $r=1$	$-2\ln LR\,(H_2(1)\vert H_3(1))$
GE/US	H_2	22.66**	6.96	0.939, 0.293, 0.003	2.52 [0.11]
JP/US	H_3	14.78*	5.40	0.928, 0.158, 0.02	3.92 [0.048]

Table 3: Cointegration tests, sample: 1985.09 - 1998.12. p-values in square brackets. $\lambda_i(\widehat{A})$: eigenvalues of the estimated companion matrix. **: significant at the 0.05, *: significant at the 0.10 levels.

	$\widehat{\beta}$ or $\widehat{\beta}^*$	$\widehat{\alpha} = \begin{pmatrix} \widehat{\alpha}_e \\ \widehat{\alpha}_p \end{pmatrix}$	$-2\ln LR\,(\gamma=-1)$	restricted $\widehat{\alpha}$
GE/US	$\begin{pmatrix} 1 \\ \widehat{\gamma} \\ \widehat{\beta}_0 \end{pmatrix} = \begin{pmatrix} 1 \\ \underset{(1.39)}{1.71} \\ \underset{(0.41)}{-0.95} \end{pmatrix}$	$\begin{pmatrix} \underset{(0.014)}{-0.047} \\ \underset{(0.0014)}{-0.0031} \end{pmatrix}$	2.18 [0.14]	$\begin{pmatrix} \underset{(0.023)}{-0.083} \\ \underset{(0.0023)}{-0.0027} \end{pmatrix}$
JP/US	$\begin{pmatrix} 1 \\ \widehat{\gamma} \end{pmatrix} = \begin{pmatrix} 1 \\ \underset{(0.61)}{-0.904} \end{pmatrix}$	$\begin{pmatrix} \underset{(0.021)}{-0.058} \\ \underset{(0.0017)}{0.0019} \end{pmatrix}$	0.01 [0.92]	$\begin{pmatrix} \underset{(0.021)}{-0.058} \\ \underset{(0.0017)}{0.0021} \end{pmatrix}$

Table 4: Parameter estimates, sample: 1985.09 - 1998.12. Standard errors in parentheses, p-values in square brackets.

to the cointegration space (henceforth model H_2).[8] The results in Table 2 support the existence of a structural break in 1985.09 affecting α and β; for this reason throughout the analysis of PPP adjustment will be based on the 'post-Plaza' period, 1985.9-1998.12.

Tables 3-4 report empirical results on PPP obtained on the sub-sample 1985.9-1998.12. Table 3 reports, for each country pair, Johansen's (1991) likelihood ratio (LR) trace tests for cointegration rank and the estimated eigenvalues of the VEC companion matrices obtained after having fixed the cointegration rank at $r = 1$; they also report the LR test for the specification H_2 against H_3 (unrestricted constant).[9] Table 4 summarizes, for each country pair, the estimated cointegrating vectors $\beta^* = (1\colon \gamma\colon \beta_0)'$ for model H_2, and $\beta = (1\colon \gamma)'$ for model H_3, the corresponding short run adjustment coefficients $\alpha = (\alpha_e\colon \alpha_p)'$, a LR test for the over-identifying restriction of long-run proportionality ($\gamma = -1$ in β, which implies that $\beta' X_t := q_t = e_t - p_t$ is mean-reverting), with the corresponding adjustment coefficients

[8] The number of lags k was fixed by combining standard information criteria with diagnostic tests on the residuals. For both country pairs we obtain $k = 2$, with insignificant residual serial correlation and moderate deviations from normality. The VEC relative to Japan and the United States includes a set of demeaned seasonal dummies, and an unrestricted impulse dummy taking values one at 1997.04, and zero elsewhere, to account for a relatively large variation in relative prices. Finally, the tests presented in Table 1 maintain that the parameters in the matrix Γ_1 of the two estimated VECs are not affected by the structural break.

[9] We also considered tests for cointegration rank r jointly with the choice of deterministic parts (model H_2 versus H_3), which consists in the joint selection procedure described in Johansen (1996), Chapter 12. This procedure ha led to the choice of models listed in Table 3 for the subsample 1985.9-1998.12. Differences in the results between models H_2 and H_3 are, in this case, negligible.

GE/US					JP/US				
A. Impact Factors									
$F_{e,q}$	$F_{p,q}$	$F_{q,\Delta e}$	$F_{q,\Delta p}$	$F_{q,q}$	$F_{e,q}$	$F_{p,q}$	$F_{q,\Delta e}$	$F_{q,\Delta p}$	$F_{q,q}$
-1.07 (0.05)	-0.07 (0.05)	12.25 (3.65)	-4.33 (13.37)	11.52 (4.18)	-0.96 (0.05)	-0.04 (0.05)	15.89 (5.43)	-25.82 (15.03)	14.57 (5.85)
B. Long-run half-lives $N_{0.5}$ (months) with 95% confidence sets ℓ_1-ℓ_2									
8	-	8	-	8	10	-	10	-	10
7-10	-	6-10	-	6-10	9-11	-	9-11	-	9-11

Table 5: Impact factors F and half-lives $N_{0.5}$, sample 1985.09-1998.12. Standard error in parenthesis.

obtained under that restriction.

Overall the results in Tables 3-4 show that long-run PPP, interpreted as stationary real exchange rates, seems to hold for both country pairs over the period 1985.09-1998.12. Even if the persistence of each estimated VEC, measured by the highest estimated eigenvalue of the companion matrix appears relatively high, in both cases the presence of I(2) stochastic trends can be ruled out.[10] Moreover, VEC estimates suggest that relative prices behave as weakly exogenous variables, i.e. they do not respond to lagged PPP deviations, whereas nominal exchange rates seem to accomplish all short-run adjustment. It can be argued, therefore, that nominal exchange rates are the primary variable that change in order to restore PPP equilibrium in the short-run. This does not necessarily mean that prices do not adjust to PPP deviations in the long-run. Long-run effects are investigated in Section 5.2.

5.2 Estimated impact factors and half-lives

Given the system $\widetilde{X}_t = (\Delta e_t : \Delta p_t : q_{t-1})'$, the IFs $F_{e,q}$ and $F_{p,q}$ quantify the 'permanent' (or 'transitory', if equal to zero) response of the nominal exchange rate e_t and the (relative) price level p_t respectively, to variations in the real exchange rate q_t. In turn, $F_{q,\Delta e}$ and $F_{q,\Delta p}$ allow to establish whether PPP deviations respond permanently (or temporarily, if equal to zero) to exchange rate depreciations and inflation differentials, respectively.

Panel A of Table 5 reports the estimated IFs for both country pairs, with corresponding standard errors. Estimates show that in both cases $F_{e,q}$ is significantly different from zero (Case 1), whereas the hypothesis $F_{p,q} = 0$ is never rejected (Case 2), suggesting that real exchange rates do not long-run Granger-cause prices, i.e. they have only transitory effects on them. Likewise, exchange rate depreciations have a permanent impact on PPP deviations ($F_{q,\Delta e} \neq 0$), whereas inflation differentials do not ($F_{q,\Delta p} = 0$).[11]

[10] We also carried out I(2) tests; the test did not imply existence of I(2) components for either country. These results are not incompatible with Bacchiocchi and Fanelli (2005), who find an I(2) stochastic trend for the GE/US pair over the longer period 1973.04-1998.02.

[11] This is further confirmed by the fact that $\widehat{F}_{q,\Delta e} \simeq \widehat{F}_{q,q}$ for both country pairs.

The results in Panel A of Table 5 remark the role of nominal exchange rates as the long-run (other than short-run) buffer of PPP deviations on the one hand, and reinforce the idea, already envisaged from the estimated adjustment coefficients αs of Table 4, that relative prices seem to behave as the stochastic common trend driving the system in the long-run, on the other hand.

The estimated IFs in Table 5 confirm that for the two country-pairs the implied half-lives $N_{0.5}(F_{e,q}(\ell), F_{e,q})$ and $N_{0.5}(F_{p,q}(\ell), F_{p,q})$ and $N_{0.5}(F_{q,\Delta e}(\ell), F_{q,\Delta e})$ and $N_{0.5}(F_{q,\Delta p}(\ell), F_{q,\Delta p})$ are not directly comparable, as argued throughout the paper. In particular, as remarked in Section 2.5, Case 2 entails 'extreme' values of the π-life, precluding a meaningful comparison with π-lives based on significant long-run effects. Panel B of Table 5 summarizes the point estimates of the 'long-run' half-lives relative to Case 1, i.e. $N_{0.5}(F_{e,q}(\ell), F_{e,q})$, $N_{0.5}(F_{q,\Delta e}(\ell), F_{q,\Delta e})$ and $N_{0.5}(F_{q,q}(\ell), F_{q,q})$, along with 95% confidence intervals which are computed following the method outlined in Section 4.2 and Appendix D. It can be noticed that the estimated $N_{0.5}$s for the nominal and real exchange rates seem in line with the prediction of sticky-price models, as the corresponding upper bound of confidence intervals do not exceed 12 months.

6 Conclusions

In this paper we address the issue of inferring the speed of adjustment of economic variables to their long-run equilibria, in the context of cointegrated VAR processes. We define the multivariate π-life as a measure of speed at which a given variable adjusts to its long-run (permanent) position, in response to variations (shocks) in another variable. The definition of π-life can be appropriately specialized, depending on whether the long-run effect is zero or not, where the latter is measured by the concept of IF. For this reason, we argue that one can hardly compare the speed of adjustment of e.g. two variables having zero and non-zero IF respectively.

The paper shows that the concept of multivariate π-life nests several special cases. For instance, when applied to interim multipliers, it delivers the π-lives of shock absorption discussed in vanDijk et al. (2007) for nonlinear systems; moreover it reduces to the traditional notion of half-life typically used by economists in the univariate framework for $\pi = \frac{1}{2}$.

We discuss likelihood-based inference on multivariate π-lives, showing that the problem of constructing confidence intervals is nonstandard. A new method is provided and its asymptotic properties discusses. It is shown how conservative confidence bounds can be obtained.

An empirical illustration focused on PPP adjustment of deutschmark-dollar and yen-dollar exchange rates reveals that the $\frac{1}{2}$-life of nominal exchange rates and prices are not directly comparable, so that one can hardly conclude that prices revert to PPP more quickly than nominal exchange rates.

Acknowledgements

Paper presented at 'Frontiers of econometric forecasting - Second workshop on dynamic econometrics in memory of Carlo Giannini', Brescia, June 9, 2006, Second Italian conference on Econometrics and Empirical Economics, Rimini, January 24-26, 2007, 'The second Tinbergen Institute conference - 20 years of cointegration: theory and practice in prospect and retrospect' Rotterdam, March 23-24, 2007, as well as in seminars in Rome, Tilburg and Oxford. We thank (without implicating) Rocco Mosconi, Riccardo Lucchetti, Martin Vagner, Chiara Osbat, Hashem Pesaran, David Hendry, Søren Johansen, Peter Boswijk, Dick vanDijk, Jurgen Doornik, Bent Nielsen, Neil Shephard and other seminar participants for useful comments on previous versions of the paper. Partial financial support is gratefully acknowledged from MIUR grants Cofin2004 and Cofin2006 (both authors), ex-60% grants University of Bologna (first author), and University of Insubria FAR 2005-2006 (second author).

References

Andrews, D.W.K, Chen, H. Y., (1994), Approximately median-unbiased estimation of autoregressive models, *Journal of Business and Economic Statistics* 12, 187-204.

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

Barro, R. J. (1990), Government spending in a simple model of endogenous growth, *Journal of Political Economy* 98, S103-S125.

Bedini, C., Mosconi, R. (2000), New tools for the dynamic analysis of cointegrated VAR models, Mimeo, Department of Economics and Production, Politecnico di Milano.

Bruneau C. and E. Jondeau (1999), Long-run causality, with application to international links between long-term interest rates, *Oxford Bulletin of Economics and Statistics* 61, 545-568.

Cheung, Y.W., Lai, K. S., (2000), On the purchasing power parity puzzle, *Journal of International Economics* 52, 321-330.

Cheung, Y.W., Lai, K. S., Bergman, M. (2004), Dissecting the PPP puzzle: the unconventional roles of nominal exchange rate and price adjustments, *Journal of International Economics* 64, 135-150.

Crowder, W.J. (2004), The converge of nominal exchange rates and price levels to the PPP equilibrium, mimeo, Department of Economics, University of Texas at Arlington, available at http://www.uta.edu/faculty/crowder/WCVITA.html.

Dornbusch, R. (1976), Expectations and exchange rate dynamics, *Journal of Political Economy* 84, 1161-1176.

Dufour J. M. and E. Renault (1998) Short tun and long run causality in time series: theory, *Econometrica* 66, 1099-1125.

Dufour J. M., D. Pelletier and E. Renault (2006) Short tun and long run causality in time series: theory, *Journal of Econometrics* 132, 337-362.

Engel, C. and Morley, J. C. (2001), The adjustment of prices and the adjustment of the exchange rate, *NBER Working Paper* W8550.

Johansen, S. (1996), *Likelihood-based inference in cointegrated Vector Auto-Regressive models*, Oxford University Press, Second revised version, Oxford.

Hansen, P. R. (2003), Structural changes in the cointegrated vector autoregressive model, *Journal of Econometrics* 114, 261-295.

Kilian, L., Zha, T. (2002), Quantifying the uncertainty about the half-life of deviations from PPP, *Journal of Applied Econometrics* 17, 107-125.

Klein, M., Mizrach B., Murphy, R.G. (1991), Managing the dollar: has the Plaza Agreement mattered?, *Journal of Money Credit and Banking* 23, 742-751.

Koop G., Pesaran M. H., Potter S. M. (1996), Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* 74, 119-147.

Lütkepohl, H. (1990) Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models, *The Review of Economics and Statistics* 72, 116–125.

Magnus J. R., H. Neudecker (1999), *Matrix differential calculus with applications in statistics and econometrics*, 2nd Edition, John Wiley and Sons, New York.

Mark, N.C. (2001), *International macroeconomics and finance; theory and econometric methods*, Blackwell Publishing.

Morley, J.C. (2007), The slow adjustment of aggregate consumption to permanent income. *Journal of Money Credit and Banking*, forthcoming.

Mussa, M. (1982), A model of exchange rate dynamics, *Journal of Political Economy* 90, 74-104.

Omtzigt, P., Paruolo, P. (2005), Impact factors, *Journal of Econometrics* 128, 31-68.

Paruolo, P. (1996), On the determination of integration indices in I(2) systems, *Journal of Econometrics* 72, 313-356.

Paruolo P. (1997), Asymptotic inference on the moving average impact matrix in cointegratared I(1) VAR systems, *Econometric Theory* 13, pp. 79-118.

Paruolo P. (2006), The likelihood ratio test for the rank of a cointegration submatrix, *Oxford Bulletin of Economics and Statistics* 68, 921-948.

Pesaran, M.H., Shin, Y. (1996), Cointegration and speed of convergence to equilibrium, *Journal of Econometrics* 71, 117-143.

Pesaran, M.H., Shin, Y. (1998), Generalized impulse response analysis in linear multivariate models, *Economic Letters* 58, 17-29.

Potter S. M. (2000) Nonlinear impulse response functions, *Journal of Economics Dynamics and Control* 24, 1425-1446.

Rogoff, K. (1996), The purchasing power parity puzzle, *Journal of Economic Literature* 34, 647-668.

Rossi, B. (2005), Confidence sets for half-life deviations from purchasing power parity, *Journal of Business and Economic Statistics* 23, 432-442.

Sims C. A. and T. Zha (1990) Error bands for impulse responses, *Econometrica* 67, 1113-1155.

vanDijk D., P. H. Franses, H. P. Boswijk (2007), Absorption of shocks in nonlinear autoregressive models, *Computational Statistics and Data Analysis*, forthcoming.

Yamamoto T. and E. Kurozumi (2006), Tests for long run Granger non-causality in cointegrated systems, *Journal of Time Series Analysis* 27, 703-723.

Appendix A: Impulse responses

In this Appendix we relate the h-step ahead multiplier $m(h)$ to impulse responses for the I(0) case. The three different measures of forecast sensitivity are (i) 'structural' impulse responses, (ii) generalized impulse responses, as defined by Koop et al. (1996) and (iii) the generalized causality-coefficients considered in Dufour and Renault (1998), Dufour et al. (2006). The relations of $m(h)$ with (i) and (ii) have already been documented in OP, section 3.5 and 6.3; here we give a few more details on these relations and add the ones with (iii). The extension of case (iii) to the I(1) case is discussed in Appendix B.

(i) Structural impulse responses

In structural impulse responses, the reduced form $p \times 1$ shocks ϵ_t are related to structural $p \times 1$ shocks η_t with a relation of the form $\epsilon_t = B\eta_t$, where B is nonsingular and typically $\text{var}(\eta_t) = I_p$, $\Omega = BB'$. The impulse response (IR) of $Y_{t+h} := J'\widetilde{X}_{t+h}$ with respect to $\eta_t = v$ is given by $\text{IR}(h, v) = J'A^h JBv$, which is seen to be a linear function of $m(h)$:

$$\text{IR}(h, v) = J'm(h)JBv.$$

(ii) Generalized impulse responses

Let Z_t be the information set available at time t, here equal to $\widetilde{X}^t_{-\infty} := (\widetilde{X}_t, \widetilde{X}_{t-1}, ...)$, and let the variables to forecast be (a subset of) $Y_{t+h} := J'\widetilde{X}_{t+h}$. The generalized impulse response (GIR) is defined as

$$\text{GIR}(h, v, Z_{t-1}) := \text{E}(Y_{t+h}|\epsilon_t = v, Z_{t-1}) - \text{E}(Y_{t+h}|Z_{t-1}) = J'A^h Jv \tag{11}$$

see Koop et al. (1996) page 133 and Pesaran and Shin (1998). Eq. (11) shows that the GIR is a linear function of the multiplier $m(h)$. For the case Gaussian errors, $\epsilon_t \sim N(0, \Omega)$, Koop et al. (1996) also consider single shocks, obtained by linear combination of ϵ_t. Let c be a $p \times 1$ selection vector and c_0 a known value; then they define the GIR with respect to $c'\epsilon_t$ as follows

$$\text{GIR}(h, c'\epsilon_t = c_0, Z_{t-1}) := \text{E}(Y_{t+h}|c'\epsilon_t = c_0, Z_{t-1}) - \text{E}(Y_{t+h}|Z_{t-1}) = J'A^h J\Omega c(c'\Omega c)^{-1}c_0. \tag{12}$$

The $\text{GIR}(h, c'\epsilon_t = c_0, Z_{t-1})$ in (12) is seen to be a special case of (11) for $v = \Omega c(c'\Omega c)^{-1}c_0$, and hence a linear function of the multiplier $m(h)$.

As a final special case, consider $c = e_i$, where e_i is the i-th column of I_p, and the choice $c_0 = (c'\Omega c)^{1/2}$. Let also GIR* be the horizontal concatenation of $\mathrm{GIR}(h, e_i'\epsilon_t = \Omega_{ii}^{1/2}, Z_{t-1})$ for $i = 1, ..., p$,

$$\mathrm{GIR}^* := (\mathrm{GIR}(h, e_1'\epsilon_t = \Omega_{11}^{1/2}, Z_{t-1}) : ... : \mathrm{GIR}(h, e_p'\epsilon_t = \Omega_{pp}^{1/2}, Z_{t-1})).$$

One finds that

$$\mathrm{GIR}^* = J'A^h J\Omega(\mathrm{dg}\,\Omega)^{-1/2} \tag{13}$$

where $\mathrm{dg}(\Omega)$ indicates a diagonal matrix of the same dimensions of Ω with the same entries of Ω on the main diagonal. (This is the expression reported for instance in OP, page 38.) This shows that also GIR* is a linear function of the multiplier $m(h)$. We note that in (12) and (13) the linear combination involves Ω, which must be estimated. The standard errors of GIR and GIR* hence require to take the estimation of Ω into account; this can be accomplished along the lines of section 6.3 in OP.

(iii) Granger causality coefficients

Dufour and Renault (1998) consider a VAR(∞), which is then simplified to a VAR(k) in Dufour et al. (2006). We take the latter formulation for compatibility with the rest of the paper, given by $X_t = \sum_{j=1}^k \Pi_j X_{t-j} + \mu_t + \epsilon_t$, where $\mu_t := \mu^* D_t^*$.

They address the forecast problem of X_{t+h} using L_2 linear projections on the closed subspaces generated by $Z_t := \{X_{t-s},\, s \geq 0\}$. The linear projection of a random variable Y on Z_t is indicated by $P(Y|Z_t)$, which coincides with the conditional expectation $\mathrm{E}(Y|Z_t)$ for linear Gaussian processes. Replacing t with $t+h$ and using recursive substitutions, one can derive the so-called '(k, h)-autoregression' representation

$$X_{t+h} = P(X_{t+h}|Z_t) + \epsilon_t^{(h)} \tag{14}$$

with $\epsilon_t^{(h)} := \sum_{j=0}^{h-1} \Pi_1^{(j)} \epsilon_{t+h-j}$ and the associated best linear predictor

$$\begin{aligned} P(X_{t+h}|Z_t) &= \sum_{j=1}^{k} \Pi_j^{(h)} X_{t+h-j} + \mu_t^{(h)} \qquad (15) \\ &= \left(\Pi_1^{(h)} : ... : \Pi_k^{(h)}\right) \begin{pmatrix} X_{t+h-1} \\ \vdots \\ X_{t+h-k} \end{pmatrix} + \mu_t^{(h)} =: \Pi^{(h)} \widetilde{X}_{t+h-1} + \mu_t^{(h)} \end{aligned}$$

where $\Pi^{(h)} := \left(\Pi_1^{(h)} : ... : \Pi_k^{(h)}\right)$, $\widetilde{X}_t := (X_{t-1}' : ... : X_{t-k+1}')'$ and

$$\Pi_j^{(s+1)} = \Pi_{j+1}^{(s)} + \Pi_1^{(s)} \Pi_j, \qquad \Pi_1^{(0)} := I_p, \qquad \Pi_1^{(1)} := \Pi_1, \qquad \mu_t^{(h)} := \sum_{j=0}^{h-1} \Pi_1^{(j)} \mu_{t+h-j}.$$

Next X_t is decomposed in 3 components, $(X_{1t}' : X_{2t}' : X_{3t}')' := (c_1 : c_2 : c_3)' X_t$ where $(c_1 : c_2 : c_3)$ is square and non singular. Dufour and Renault define

$$\mathrm{GC}_{c_1,c_2}(h, j) := c_1' \Pi_j^{(h)} c_2, \qquad j = 1, 2, .., k, \quad h = 1, 2, ...$$

as the 'generalized impulse response coefficients' of X_{1t+h} with respect to changes in X_{2t+h-j}. These coefficients "provide a complete picture of the linear causality properties at different horizons", see Dufour and Renault (1998), page 1113. For a given forecast horizon h, one can group all the $\mathrm{GC}_{c_1,c_2}(h,j)$ into matrix as follows:

$$\mathrm{GC}_{c_1,c_2}(h) := (\mathrm{GC}_{c_1,c_2}(h,1) : ... : \mathrm{GC}_{c_1,c_2}(h,k)) = c_1'\Pi^{(h)}(I_k \otimes c_2).$$

They prove that the restrictions $\mathrm{GC}_{c_1,c_2}(h) = 0$ is a necessary and sufficient condition for X_{2t} not to Granger-cause X_{1t} at forecast horizon h when the variance covariance matrix of $\epsilon_t^{(h)}$ is nonsingular, and that without the last proviso the condition $\mathrm{GC}_{c_1,c_2}(h) = 0$ is sufficient. Note that this is a linear restriction on $\Pi^{(h)}$.

Here we show that $\Pi^{(h)}$ is a linear function of $m(h)$, and hence a fortiori, that also $\mathrm{GC}_{c_1,c_2}(h)$ is a linear function of $m(h)$. To this end one can for instance employ the state space dynamics (2) to derive $P(\widetilde{X}_{t+h}|Z_t) = A^h\widetilde{X}_{t+h+1} + \sum_{i=0}^{h-1} A^iJ\mu_{t+h-i}$. Because $X_t = J'\widetilde{X}_t$, one finds $P(X_{t+h}|Z_t) = J'P(\widetilde{X}_{t+h}|Z_t) = J'A^h\widetilde{X}_{t+h+1} + \sum_{i=0}^{h-1} J'A^iJ\mu_{t+h-i}$. Equating coefficients one finds $\Pi^{(h)} = J'A^h = J'm(h)$, i.e. that $\Pi^{(h)}$ and hence $\mathrm{GC}_{c_1,c_2}(h)$ are linear functions of $m(h)$.

Appendix B: Long-run Granger causality

In this Appendix we describe how the conditions for long-run noncausality for I(1) systems as defined in Bruneau and Jondeau (1999) and Yamamoto and Kurozumi (2006) can be expressed as hypothesis on the IF F. Hence tests of long-run noncausality or neutrality can be considered as special cases of tests on the impact factors F.

Long-run Granger noncausality is defined by the above authors in terms of $X_{\infty|t}$, see (8), as follows. Consider some target variables $b_1'X_t$ and some candidate causal variables $a_1'X_t$, where a_1, b_1 are of dimension $p \times n$ and $p \times m$ respectively and $(a_1 : b_1)$ is of full column rank. Then $a_1'X_t$ is said not to Granger cause $b_1'X_t$ in the long run if $b_1'X_{\infty|t}$ does not depend on $(X_t'a_1 : X_{t-1}'a_1 : ... : X_{t-k+1}'a_1)'$, which contains $a_1'X_t$ and its lags. Bruneau and Jondeau (1999) show that this corresponds to the conditions

$$b_1'Ca_1 = 0, \tag{16}$$

$$b_1'C\Gamma_i a_1 = 0, \qquad i = 1, ..., k-1. \tag{17}$$

Condition (16) is also called 'long-run neutrality' by Yamamoto and Kurozumi (2006), see their definition 2. The conditions (16) and (17) can be phrased as restrictions on the first block of p rows in the impact factors F in Section 3 above.

Note that tests of condition (16) are discussed in Paruolo (1997) when $n = m = 1$, see also Paruolo (2006). Tests of (16), (17) are also considered in Bruneau and Jondeau (1999) and Yamamoto and Kurozumi (2006). These tests are relevant in the setup of this paper in order to empirically distinguish between Case 1 and Case 2.

Appendix C: Proofs

In this appendix we report proofs of propositions in Section 4.2.

Proof. of Proposition 1. Let $h^{-1}(\mathcal{H}_\pi)$ be the inverse image of $\mathcal{H}_\pi$. It is simple to see that $\mathcal{A} \subset h^{-1}(\mathcal{H}_\pi)$, so that $\Pr(h \in \mathcal{H}_\pi) \geq \Pr(A \in \mathcal{A}) = 1 - \eta$. ■

In order to prove Proposition 2, we introduce the following notation. Let $(A_{\ell,1}, A_{\ell,2})$ equal $(A_{\ell,\min}, A_{\ell,\max})$ if $|\varphi_{\ell,\min}| \leq |\varphi_{\ell,\max}|$ and equal $(A_{\ell,\max}, A_{\ell,\min})$ otherwise. Define also for $j = 1, 2$

$$\mathcal{A}_j := \{A_{\ell,j}, \ell \in \mathbb{N}\},$$

where $\mathcal{A}_j \subset \mathcal{A}$. We note that $\ell_2 = \min \ell : |\varphi_\ell(A_{\ell,2})| \leq \pi$ and $\ell_1 = \min \ell : |\varphi_\ell(A_{\ell,1})| \leq \pi$.

Proof. of Proposition 2. Let $A^* \in \mathcal{A}$ correspond to $\ell^* := h^\pi_{\max} = h_\pi(A^*)$. By definition $\pi < |\varphi_{\ell^*-1}(A^*)|$ and $|\varphi_s(A^*)| \leq \pi$ for $s \geq \ell^*$. Take now $A_{\ell^*-1,2}$ which by definition satisfies $\pi < |\varphi_{\ell^*-1}(A^*)| \leq |\varphi_{\ell^*-1}(A_{\ell^*-1,2})|$. Now consider $|\varphi_s(A_{\ell^*-1,2})|$ for $s \geq \ell^*$. If $\pi < |\varphi_s(A_{\ell^*-1,2})|$ for some $s \geq \ell^*$, this would imply a contradiction to the assumption $\ell^* := h^\pi_{\max} = h_\pi(A^*)$. Hence it must hold that $|\varphi_s(A_{\ell^*-1,2})| \leq \pi$ for all $s \geq \ell^*$, i.e. $h_\pi(A_{\ell^*-1,2}) = h_\pi(A^*) = \ell^* := h^\pi_{\max}$. Because $\mathcal{A}_2 \subset \mathcal{A}$, one has $\max_{A\in\mathcal{A}} h_\pi(A) \geq \max_{A\in\mathcal{A}_2} h_\pi(A)$. Hence we have shown that $\max_{A\in\mathcal{A}} h_\pi(A) = h_\pi(A_{\ell^*-1,2}) = \ell_2$.

Consider next $h^\pi_{\min}$. By definition $\pi < |\varphi_s(A_{s,1})|$ for $s < \ell_1$. Because of the extreme properties of $A_{s,1}$, see problems in (10), one has $\pi < |\varphi_s(A_{s,1})| \leq |\varphi_s(A)|$ for all $A \in \mathcal{A}$, $s < \ell_1$, and hence $h^\pi_{\min} \geq \ell_1$. ■

Appendix D: Optimization

We here discuss optimization of

$$\varphi_\ell = \frac{F_{b,a}(\ell)}{F_{b,a}} - 1 = -\frac{b' A^{\ell+1} (I - A)^{-1} a}{b' \left((I - A)^{-1} - I\right) a} =: \frac{c_1}{c_2}.$$

In the following Proposition 3 we state first and second derivatives of φ_ℓ as a function of $x := \operatorname{vec}(A')$. We observe that the parameters that vary in $\mathcal{A}$ are $\xi := \operatorname{vec}(A'J) = H'x$, with $H' := (J' \otimes I_p) = (I_g : 0)$. Let $\mathcal{X}$ be the set of values of ξ that correspond to $A \in \mathcal{A}$. One can hence use the chain rule of derivatives to compute

$$\dot{\varphi}_\ell(\xi) := \frac{\partial \varphi_\ell(\xi)}{\partial \xi'} = \frac{\partial \varphi_\ell(x)}{\partial x'} H, \qquad \ddot{\varphi}_\ell(\xi) := \frac{\partial^2 \varphi_\ell(\xi)}{\partial \xi \partial \xi'} = H' \frac{\partial^2 \varphi_\ell(x)}{\partial x \partial x'} H.$$

As in Newton-like methods, we consider a second order approximation f of φ_ℓ around a given value ξ_0 of ξ

$$f(\xi) := \varphi_0 + \dot{\varphi}_0 (\xi - \xi_0) + \frac{1}{2} (\xi - \xi_0)' \ddot{\varphi}_0 (\xi - \xi_0).$$

where $\varphi_0 := \varphi(\xi_0)$, $\dot{\varphi}_0 := \dot{\varphi}(\xi_0)$ and $\ddot{\varphi}_0 := \ddot{\varphi}(\xi_0)$. Unlike standard least-square problems, $f(\xi)$ is bound to be non-concave (non-convex) as a function of ξ also in the proximity of $\arg\max_{\xi \in \mathcal{X}} f(\xi)$ (or $\arg\min_{\xi \in \mathcal{X}} f(\xi)$); this may cause convergence problems to standard quasi-Newton methods. We hence introduce a Newton-like algorithm suitable for the current situation.

Consider the eigenvalues of $\ddot{\varphi}_0$, partitioned into the negative and positive ones, $\lambda_1 \leq ... \leq \lambda_{g_1} < 0 < \lambda_{g_1+1} \leq ... \leq \lambda_g$. Define $\Lambda_1 := -D_1 :=$diag$(\lambda_1 : ... : \lambda_{g_1})$ and $\Lambda_2 := D_2 :=$diag$(\lambda_{g_1+1} : ... : \lambda_g)$, where D_i are positive definite (p.d.) by construction for $i = 1, 2$. Partition the eigenvectors W conformably with D_1 and D_2, so that one has the spectral decomposition

$$\ddot{\varphi}_0 = W\Lambda W' = (W_1 : W_2)\,\mathrm{diag}(\Lambda_1, \Lambda_2)\,(W_1 : W_2)' = -W_1 D_1 W_1' + W_2 D_2 W_2'.$$

(If some eigenvalues of $\ddot{\varphi}_0$ are 0, then they are simply omitted in the spectral decomposition above, which is still valid). We here show that when one selects

$$\xi_i := \xi_0 - W_i \Lambda_i^{-1} W_i' \dot{\varphi}_0, \qquad i = 1, 2,$$

one obtains $f(\xi_1) > f(\xi_0)$ and $f(\xi_2) < f(\xi_0)$ respectively; this ensures a step in the right direction for the optimization problems $\max f(\xi)$ and $\min f(\xi)$. In fact substituting into $f(\xi)$ one finds

$$f(\xi_i) - f(\xi_0) = -\dot{\varphi}_0 W_i \Lambda_i^{-1} W_i' \dot{\varphi}_0 + \frac{1}{2}\dot{\varphi}_0 W_i \Lambda_i^{-1} W_i' \dot{\varphi}_0 = -\frac{1}{2}\dot{\varphi}_0 W_i \Lambda_i^{-1} W_i' \dot{\varphi}_0$$

which imply

$$\begin{aligned} f(\xi_1) - f(\xi_0) &= \frac{1}{2}\dot{\varphi}_0 W_1 D_1^{-1} W_1' \dot{\varphi}_0 > 0 \\ f(\xi_2) - f(\xi_0) &= -\frac{1}{2}\dot{\varphi}_0 W_2 D_2^{-1} W_2' \dot{\varphi}_0 < 0 \end{aligned}$$

due to the fact that D_i are p.d. Hence this defines a Newton-like algorithm that selects ξ_1 as direction vector for $\max f(\xi)$ and ξ_2 as direction vector for $\min f(\xi)$. We modify this Newton-type algorithm including a line search in direction ξ_i.

We finally report first and second derivatives of $\varphi_\ell(x)$ in the following proposition. Denote by $K := F + I$, and let also $\mathcal{K}$ be the commutation matrix that satisfies $\mathrm{vec}(A) = \mathcal{K}\,\mathrm{vec}(A')$, see e.g. Magnus and Neudecker (1999).

Proposition 3 *$\varphi_\ell(x)$ is continuously differentiable with gradient*

$$\frac{\partial \varphi_\ell(x)}{\partial x'} = -\frac{1}{c_2}\left(b' \otimes a'K'\right)\left(Q_\ell + \left(A^{\ell+1}K \otimes I\right) + \varphi_\ell(K \otimes I)\right)$$

and Hessian

$$\frac{\partial \varphi_\ell}{\partial x \partial x'} = \frac{1}{c_2}\left(\mathcal{G} + 2\frac{c_1}{c_2^2}\mathcal{K}\left((K \otimes K'ba'K') + (Kab'K \otimes K')\right)\right) - \frac{1}{c_2^2}\left(R + R'\right)$$

where $R := (Kbb' \otimes Kaa'K)\left(Q_\ell + \left(A^{\ell+1}K \otimes I\right)\right)$, $Q_\ell := \sum_{i=0}^{\ell} A^i \otimes \left(A^{\ell-i}\right)'$ *and*

$$\begin{aligned} \mathcal{G} &= -\sum_{i=0}^{\ell-1}\sum_{j=0}^{\ell-i-1}\mathcal{K}\left(A^j \otimes (A^i)' ba'K'\left(A^{\ell-i-1-j}\right)'\right) - \sum_{i=1}^{\ell}\sum_{j=0}^{i-1}\mathcal{K}\left(A^{\ell-i}Kab'A^j \otimes \left(A^{i-1-j}\right)'\right) \\ &\quad -\mathcal{K}\left(Kab'A^{\ell+1}K \otimes K'\right) - \sum_{j=0}^{\ell}\mathcal{K}\left(Kab'A^j \otimes K'\left(A^{\ell-j}\right)'\right). \end{aligned}$$

When $\ell \to \infty$ one has

$$\begin{aligned} \frac{\partial \varphi_\ell(x)}{\partial x'} &\to -\frac{\varphi_\ell}{c_2}\left(b'K \otimes a'K'\right) \\ \frac{\partial \varphi_\ell}{\partial x \partial x'} &\to 2\frac{c_1}{c_2^3}\mathcal{K}\left(\left(K \otimes K'ba'K'\right) + \left(K'ba'K' \otimes K\right)\right) \end{aligned}$$

Proof. Differentiating φ_ℓ one finds $\mathrm{d}\varphi_\ell = \mathrm{d}c_1 c_2^{-1} - c_1 c_2^{-2}\mathrm{d}c_2$, where

$$\begin{aligned} \mathrm{d}c_1 &= -a'K'\left(\sum_{i=0}^{\ell}\left(A^{\ell-i}\right)'(\mathrm{d}A)'\left(A^i\right)' + (\mathrm{d}A)'K'\left(A^{\ell+1}\right)'\right)b, \\ \mathrm{d}c_2 &= b'K(\mathrm{d}A)Ka = a'K'(\mathrm{d}A)'K'b. \end{aligned}$$

Hence one has

$$\begin{aligned} &\frac{\partial c_1}{\partial x'} = -\left(b' \otimes a'K'\right)\left(Q_\ell + A^{\ell+1}K \otimes I\right), \qquad \frac{\partial c_2}{\partial x'} = b'K \otimes a'K' \\ &\frac{\partial \varphi_\ell(x)}{\partial x'} = \frac{1}{c_2}\left(\frac{\partial c_1}{\partial x'} - \varphi_\ell\frac{\partial c_2}{\partial x'}\right) = -\frac{1}{c_2}\left(b' \otimes a'K'\right)\left(Q_\ell + A^{\ell+1}K \otimes I + \varphi_\ell(K \otimes I)\right) \end{aligned}$$

The term $Q_{\ell+1}$ satisfies the recursion $Q_{\ell+1} = (A \otimes I)Q_\ell + \left(I \otimes \left(A^{\ell+1}\right)'\right)$ in ℓ starting from $Q_0 = I$. This can be proved by observing that

$$\begin{aligned} Q_{\ell+1} &= \sum_{i=0}^{\ell+1} A^i \otimes \left(A^{\ell+1-i}\right)' = (A \otimes I)\sum_{i=1}^{\ell+1} A^{i-1} \otimes \left(A^{\ell-(i-1)}\right)' + I \otimes \left(A^{\ell+1}\right)' \\ &= (A \otimes I)Q_\ell + I \otimes \left(A^{\ell+1}\right)'. \end{aligned}$$

This recursions in ℓ also shows that Q_ℓ tends to 0 as $\ell \to \infty$, because A is stable. For the same reason also $A^{\ell+1}K \otimes I$ and the whole $\partial c_1/\partial x'$ tends to 0 as $\ell \to \infty$. This implies the limit behavior of the first derivative for $\ell \to \infty$. We next consider the second differential of φ_ℓ in direction x and $\widetilde{x}$; we use $\widetilde{}$ to indicate increments in the second direction

$$\begin{aligned} \mathrm{d}^2\varphi_\ell(\mathrm{d}x, \mathrm{d}\widetilde{x}) &= \mathrm{d}^2 c_1 c_2^{-1} - c_2^{-2}\mathrm{d}c_1\mathrm{d}\widetilde{c}_2 - \mathrm{d}\widetilde{c}_1 c_2^{-2}\mathrm{d}c_2 + 2c_1 c_2^{-3}\mathrm{d}^2 c_2 \\ &= c_2^{-1}\left(\mathrm{d}^2 c_1 + 2c_1 c_2^{-2}\mathrm{d}^2 c_2\right) - c_2^{-2}\left(\mathrm{d}c_1\mathrm{d}\widetilde{c}_2 + \mathrm{d}\widetilde{c}_1\mathrm{d}c_2\right), \\ \mathrm{d}^2 c_1(\mathrm{d}x, \mathrm{d}\widetilde{x}) &= -a'K'\left(\sum_{i=0}^{\ell-1}\widetilde{\mathrm{d}}\left(A^{\ell-i}\right)'(\mathrm{d}A)'\left(A^i\right)' + \sum_{i=1}^{\ell}\left(A^{\ell-i}\right)'(\mathrm{d}A)'\widetilde{\mathrm{d}}\left(A^i\right)'\right)b \\ &\quad - a'K'\left((\mathrm{d}A)'\widetilde{\mathrm{d}}K'\left(A^{\ell+1}\right)' + (\mathrm{d}A)'K'\widetilde{\mathrm{d}}\left(A^{\ell+1}\right)'\right)b =: c_3 + c_4 + c_5 + c_6 \\ \mathrm{d}^2 c_2(\mathrm{d}x, \mathrm{d}\widetilde{x}) &= \mathrm{tr}\left(K'ba'K'\mathrm{d}\widetilde{A}'K'\mathrm{d}A' + K'\mathrm{d}\widetilde{A}'K'ba'K'\mathrm{d}A'\right) = \\ &= \mathrm{d}x'\mathcal{K}\left(\left(K \otimes K'ba'K'\right) + \left(Kab'K \otimes K'\right)\right)\mathrm{d}\widetilde{x} \end{aligned}$$

where

$$
\begin{aligned}
c_3 &= -a'K'\left(\sum_{i=0}^{\ell-1}\left(\sum_{j=0}^{\ell-i-1} A^j \mathrm{d}\widetilde{A} A^{\ell-i-1-j}\right)' \mathrm{d}A' \left(A^i\right)'\right) b \\
&= -\sum_{i=0}^{\ell-1}\sum_{j=0}^{\ell-i-1} \operatorname{tr}\left(\left(A^i\right)' ba'K' \left(A^{\ell-i-1-j}\right)' \mathrm{d}\widetilde{A}' \left(A^j\right)' \mathrm{d}A'\right) \\
&= -\sum_{i=0}^{\ell-1}\sum_{j=0}^{\ell-i-1} \mathrm{d}x'\mathcal{K}\left(A^j \otimes \left(A^i\right)' ba'K' \left(A^{\ell-i-1-j}\right)'\right) \mathrm{d}\widetilde{x} =: \mathrm{d}x' C_3 \mathrm{d}\widetilde{x}
\end{aligned}
$$

$$
\begin{aligned}
c_4 &= -\sum_{i=1}^{\ell}\sum_{j=0}^{i-1} \operatorname{tr}\left(\left(A^{i-1-j}\right)' \mathrm{d}\widetilde{A}' \left(A^j\right)' ba'K' \left(A^{\ell-i}\right)' \mathrm{d}A'\right) \\
&= -\sum_{i=1}^{\ell}\sum_{j=0}^{i-1} \mathrm{d}x'\mathcal{K}\left(A^{\ell-i}Kab'A^j \otimes \left(A^{i-1-j}\right)'\right) \mathrm{d}\widetilde{x} =: \mathrm{d}x' C_4 \mathrm{d}\widetilde{x}
\end{aligned}
$$

$$
\begin{aligned}
c_5 &= -a'K' \left(\mathrm{d}A\right)' K' \left(\mathrm{d}\widetilde{A}\right)' K' \left(A^{\ell+1}\right)' b = -\operatorname{tr}\left(K' \left(\mathrm{d}\widetilde{A}\right)' K' \left(A^{\ell+1}\right)' ba'K' \left(\mathrm{d}A\right)'\right) \\
&= -\mathrm{d}x'\mathcal{K}\left(Kab'A^{\ell+1}K \otimes K'\right) \mathrm{d}\widetilde{x} =: \mathrm{d}x' C_5 \mathrm{d}\widetilde{x}
\end{aligned}
$$

$$
\begin{aligned}
c_6 &= -a'K' \left(\left(\mathrm{d}A\right)' K' \widetilde{\mathrm{d}} \left(A^{\ell+1}\right)'\right) b = -a'K' \left(\mathrm{d}A\right)' K' \left(\sum_{j=0}^{\ell} A^j \mathrm{d}\widetilde{A} A^{\ell-j}\right)' b \\
&= -\sum_{j=0}^{\ell} \operatorname{tr}\left(K' \left(A^{\ell-j}\right)' \mathrm{d}\widetilde{A}' \left(A^j\right)' ba'K' \mathrm{d}A'\right) = -\sum_{j=0}^{\ell} \mathrm{d}x'\mathcal{K}\left(Kab'A^j \otimes K' \left(A^{\ell-j}\right)'\right) \mathrm{d}\widetilde{x} \\
&=: \mathrm{d}x' C_6 \mathrm{d}\widetilde{x}
\end{aligned}
$$

Collecting terms and setting $\mathcal{G} := \sum_{i=3}^{6} C_i$, one finds the expression of the Hessian given above. ■