WP 25_12

Karim M. Abadir
Imperial College London, UK
The Rimini Centre for Economic Analysis (RCEA), Italy

Gabriel Talmain
University of Glasgow, UK

BEYOND CO-INTEGRATION: MODELLING CO-MOVEMENTS IN MACRO FINANCE

Beyond Co-Integration: Modelling Co-Movements in Macro Finance[*]

KARIM M. ABADIR[†] and GABRIEL TALMAIN[‡]

[†]Tanaka Business School, Imperial College London, London SW7 2AZ, UK, e-mail: k.m.abadir@imperial.ac.uk

[‡]Department of Economics, Adam Smith Building, 40 Bute Gardens, University of Glasgow, Glasgow G12 8RT, UK, e-mail: g.talmain@lbss.gla.ac.uk

Beyond Co-Integration: Modelling Co-Movements in Macro Finance

ABSTRACT

Macroeconomic and aggregate financial series share an unconventional type of nonlinear dynamics. Existing techniques (like co-integration) model these dynamics incompletely, hence generating seemingly paradoxical results. To avoid this, we provide a methodology to disentangle the long-run relation between variables from their own dynamics, and illustrate with two applications.

First, in the forward-premium puzzle, adding a component quantifying the persistent nonlinear dynamics of exchange rates yields substantial predictability and makes the forward-premium term insignificant. Second, S&P 500 grows in a pattern of momentum followed by reversal, forming long cycles around a trend given by GDP, a stable non-breaking relation since WWII.

There are a number of paradoxes or puzzles, instances in which some very natural relation between economic or financial variables seems to be violated, with the estimated coefficients of the relation defying financial or economic logic. Many of these paradoxes concern cases where the dynamics of the variables are persistent and not easily modelled empirically. We will illustrate how these dynamics are responsible for such puzzles, and we will present new tools for handling relations between persistent series in a way that avoids spurious correlations between these variables.[1]

The dynamics of a series can be depicted by its Auto-Correlation Function (ACF), in addition to its usual time-domain and frequency-domain representations. Figures 1 and 2 show the ACFs of the logarithms of the $/£ exchange rate and the real S&P 500 stock market index, the formula for the fitted curve in each of these graphs to be defined later in (9). For the moment, one can immediately observe how similar the shape of these two functions are, and how strong the autocorrelation is, even at long lags. These well-structured and smooth ACFs are striking to anyone used to seeing the jagged time-paths of these variables. It is also striking how different these ACFs are from the ones implied by Auto-Regressive Integrated Moving-Average (ARIMA) models that include the special cases of unit roots and random walks.

If the dynamics of the variables in a model are not adequately represented, biased and inconsistent estimates of the relation will arise. The current thinking is that the theory of integration and co-integration can deal with such issues and avoid spurious correlations, and the theory has developed into a huge branch of econometrics to try to deal with series having a high degree of persistence. A defining feature of integration is that it assumes that the dynamics of individual series can be approximated parsimoniously by a class of linear processes which, for example, Figures 1 and 2 reveal not to be the case. This is why a new econometric methodology is

3

needed to deal with estimating relations between variables containing this type of nonlinear long memory.[2] Our proposed approach will encompass single-equation co-integration models as a special case.

We exploit the common structure of these ACFs to devise a simple new method to disentangle the co-movements of variables (i.e. identify and estimate the parameters of the relation linking them) from the effects of persistence of the individual series. We provide two applications of different natures to illustrate the versatility and potential of our method. These applications should not be misconstrued as complete modelling exercises, as such additional endeavours would be far too substantial to fit in a single study. The first application shows how our method dramatically reverses the counterintuitive finding about the forward premium puzzle in the Uncovered Interest Parity (UIP) regressions, revealing the insignificance of the premium's role and the importance of the predictable own-dynamics of exchange rates. The second application is about significance of a relation between two variables, rather than the lack of it. It shows how the stock market grows in long cycles around a trend given by Gross Domestic Product (GDP), in a stable relation that does not break and that fits better than existing econometric methods: S&P's persistent dynamics and GDP alone explain more than 50% of the variation in stock *returns* during the period 1953–2004, which is an exceptionally high R^2 given that we have not used any other variables.[3] Inter alia, our model also predicts the rebound that followed the overreaction to the oil crisis in 1973, and the bursting of the dot-com bubble of the late 1990's. Using the methodology of Campbell and Thompson (2008), we also show that our model could help a risk-averse investor to reduce the weight of the risk-free asset in her optimal portfolio and produce large gains in the portfolio's expected return.

Our two applications encompass parsimoniously many salient features obtained in a growing body of evidence for the predictability of exchange

rates and stock markets. We parameterize these persistent nonlinear dynamics in one common formulation that is consistent with the microfounded general-equilibrium model of Abadir and Talmain (2002) and with the evidence uncovered here. In the case of exchange rates, strong autocorrelations have been stressed in Backus, Gregory, and Telmer (1993), Bekaert (1996), Okunev and White (2003); while Engel and Hamilton (1990), Diebold, Husted, and Rush (1991), Diebold, Gardeazabal, and Yilmaz (1994), Baillie and Bollerslev (1994, 2000), Sarno, Valente, and Leon (2006) have presented evidence of long swings and persistence. For the stock market, in the short and intermediate terms, the existence of momentum for individual stocks has been shown by Jegadeesh and Titman (1993, 2001), Grundy and Martin (2001), Grinblatt and Moskowitz (2004), Hvidkjaer (2006). The subsequent reversal of this momentum into a correction or cycle has been documented in De Bondt and Thaler (1985, 1987, 1989), Lakonishok, Shleifer, and Vishny (1994), Lee and Swaminathan (2000), Griffin, Ji, and Martin (2003), Cooper, Gutierrez, and Hameed (2004), Bhojraj and Swaminathan (2006). The empirical relation of the stock market to the underlying state of the economy and the implied predictability have been discussed in Campbell and Shiller (1988) Fama and French (1989), and more recently in Liew and Vassalou (2000), Lettau and Ludvigson (2001), Vassalou (2003), Chordia and Shivakumar (2006), Petkova (2006), Cochrane (2008), Campbell and Thompson (2008).

The plan of our paper is as follows. Section I presents a baseline version of the forward premium puzzle, showing how the residual diagnostics indicate omitted nonlinear long memory. Section II outlines our procedure and explains the essentials of how it deals with the residual's unconventional dynamics. Sections III and IV apply our procedure to the forward premium puzzle and stock market, respectively. Section V concludes. The Appendix gives the technical details relating to our procedure. We follow the notation

conventions proposed in Abadir and Magnus (2002).

I. The UIP Theorem and the Forward Premium Puzzle

One test for the efficiency of the foreign exchange market, going back to Fisher (1930), is that "speculators" will equalize the expected return on the similar short term assets across countries once converted to the same currency. However, a large number of authors analyzing the data have found systematic deviations from this norm. The data seemed to lend support to a very substantial negative relation between the returns on holding a currency and the forward premium on it. This is known as the forward premium puzzle or anomaly. Many authors have studied this very counterintuitive result and excellent summaries are found in Froot and Thaler (1990), Lewis (1995), Engel (1996).

This section contains two parts: the three alternative formulations of the baseline UIP theorem, followed by the empirical puzzle to be illustrated graphically and by regression. The first formulation, to be given in (2), is in terms of excess returns and sets the stage for a simple graphical presentation. The second, in (3), is in terms of depreciation of a currency, and is widely used in the literature. It also provides a bridge to the third form, in (4), which is in terms of the levels of the variables and shows how the UIP regression can be expressed in terms of our estimation method. Note that Apte, Sercu, and Uppal (2004) recommend using levels in the related context of purchasing power parity.

A. Three Forms of the Baseline UIP Regression

Consider a US investor who, at time t, can invest either in a domestic dollar-denominated bond or in a foreign sterling-denominated bond. The nominal interest rate, paid at maturity $t + 1$, is I_t for the domestic bond and I_t^* for

the foreign bond. The interest rates I_t and I_t^* are agreed upon and known at time t. Let the exchange rate be S_t, such that one pound sterling is worth S_t US dollars. Consider the following two alternatives, each involving an investment of \$1 at time t:

1. Invest in the domestic US\$ bond. The yield at time $t+1$ is \$$I_t$, which can be approximated by $i_t \equiv \log(1 + I_t)$ since $\log(1 + I_t) \approx I_t$ for small I_t.

2. Invest in the foreign UK£ bond. The \$1 is first converted at the current exchange rate into £$1/S_t$. This amount is then invested in the foreign bond at time t, to produce £$(1 + I_t^*)/S_t$ at time $t+1$. Converting it back into dollars at the new exchange rate S_{t+1} gives \$$(1 + I_t^*)S_{t+1}/S_t$ and we subtract 1 to get the yield on this investment. Using $a - 1 \approx \log a$, the approximate US\$ yield is $i_t^* + s_{t+1} - s_t$, where $i_t^* \equiv \log(1 + I_t^*)$ and $s_t \equiv \log S_t$. The difference $\Delta s_{t+1} \equiv s_{t+1} - s_t$ is the approximate rate of depreciation of the US currency.

Ignoring transaction costs, the excess return on investing in the foreign asset is then defined as

$$r_{t+1} \equiv i_t^* - i_t + \Delta s_{t+1}. \tag{1}$$

The UIP hypothesis implies that r_{t+1} should not be predictable. In particular, the forward premium $(f_t - s_t)$, where f_t is the log of the forward rate F_t, should have no explanatory power. We briefly consider three essentially equivalent formulations of a test for this latter hypothesis.

The first form is a direct implementation, estimating

$$r_{t+1} = \alpha + \beta(f_t - s_t) + u_{t+1} \tag{2}$$

and checking if $\beta = 0$. The puzzle is that the literature has found significantly negative estimates for β. A higher premium $(f_t - s_t)$ means that the

forward rate indicates that the US\$ is likely to depreciate, and yet the regression says that it is systematically associated instead with positive excess returns being made on the US\$ (i.e. lower r_{t+1} on the UK£). The idea that market participants are ready to pay more for an asset when they expect it to become less attractive seem to fly in the face of market efficiency or even of rationality. A positive β might have been excusable, but a very large negative one is puzzling.

The second form of the UIP regression can be derived by using the Covered Interest Parity (CIP) relation. Consider another alternative for investing \$1, which is to convert it into £$1/S_t$, invest this amount in the foreign bond and sell forward the forthcoming £$(1 + I_t^*)/S_t$ at the forward rate F_t. Since all of these transactions can be completed today at no risk, the US\$ yield on this, $i_t^* + f_t - s_t$, must be equal to the US\$ yield i_t of investing in a domestic bond, by arbitrage. Hence, $i_t^* - i_t = -(f_t - s_t)$. By substituting this into (2) and using definition (1), we get the second form of the UIP regression:

$$\Delta s_{t+1} = \alpha + (1 + \beta)(f_t - s_t) + u_{t+1}. \tag{3}$$

Formulations (2) and (3) are equivalent, up to the CIP relation. Unlike the UIP relation, one can verify that the CIP holds almost exactly in the major markets (hence $r_{t+1} \approx s_{t+1} - f_t$ from (1)). Note that $\beta = -1$ would imply that the exchange rate follows a random walk if u_{t+1} did not contain further dynamics, but this random walk hypothesis for s_t is negated visually by Figure 1. A more thorough testing of β will follow in Section III. A random walk can arise alternatively if $\{f_t - s_t\}$ is i.i.d. or, as a degenerate special case of it, $\{f_t - s_t\}$ is approximately constant (translating into an additional drift for the $\{s_t\}$ process). The latter assumption is used in pricing currency options.

The third form of the UIP regression is obtained by recalling that $\Delta s_{t+1} \equiv$

$s_{t+1} - s_t$ and adding s_t to both sides of (3):

$$s_{t+1} = \alpha + \gamma f_t - \beta s_t + u_{t+1}, \qquad (4)$$

where $\gamma \equiv 1 + \beta$, the test becoming $\beta = 0$ and $\gamma = 1$. This formulation is in terms of the levels of the variables, with s_{t+1} as the dependent variable and only u_{t+1} is contemporaneous to it in the equation.

B. The Forward Premium Puzzle

We start by presenting the results using traditional methods, to verify the presence of the puzzle in monthly data from the Bank of International Settlements (BIS). We obtained data for the period January 1979 to February 2004. Running the regression (4) on the original data, with all coefficients unrestricted, Ordinary Least Squares (OLS) gives

$$
\begin{aligned}
\widehat{s}_{t+1} \quad = \quad &-2.65\, f_t \quad +3.63\, s_t, \\
&(-5.53) \quad\;\; (5.50) \\
&[-3.69] \quad\;\; [3.66]
\end{aligned}
\qquad (5)
$$

where we do not report the estimated constant because it is tangential to the analysis of predictability. The t-ratios and their Heteroskedastic and Autocorrelation-Consistent (HAC) versions are given in parentheses and brackets, respectively. The hypotheses they test are that the coefficients of f and s equal 1 and 0, respectively.

There is a substantial amount of autocorrelation and heteroskedasticity left over in the residuals, as is evidenced by the difference between the adjusted and unadjusted t-ratios. Nevertheless, inference that is robust to standard dynamics in the residuals can be carried out using HAC t-ratios. They show that the coefficients of f and s are significantly different from their anticipated values of 1 and 0, respectively.

It seems that forward rates violate the UIP in a puzzling way, if one were to believe these estimated parameters. Based on (2), a scatter plot

9

of the excess return r_{t+1} (on the vertical axis) versus the forward premium $(f_t - s_t)$ (on the horizontal axis) in Figure 3 tells the story. The data form a funnel shape (indicating heteroskedasticity), with a clear negative inclination $(\beta < 0)$. As we shall show, these distortions are due to the graph (or regression) missing a third dimension: the nonlinear long-memory dynamics.

Froot and Thaler (1990, p.188), put their finger on the problem, noting that it could be explained "if only part of this appreciation occurs immediately, and the rest takes some time". We will show that this is precisely the sort of problem that current econometric techniques cannot adequately deal with. We now provide the required technique, and show how it can solve the puzzle.

II. The ACF-Based Procedure

Consider the general representation

$$z_t = \widetilde{z}_t + u_t, \quad t = 1, 2, \ldots T, \tag{6}$$

where $\widetilde{z}_t$ represents the time-varying 'fundamental' value of z_t, while u_t are the residual dynamics of adjustment towards such a value. By definition, u_t is centered around zero and is mean-reverting, otherwise z_t will not revert to its fundamental value $\widetilde{z}_t$. Denoting the $T \times 1$ vector of stacked z_t values by $\boldsymbol{z} \equiv (z_1, \ldots, z_T)'$, and similarly for $\widetilde{z}_t$ and u_t, we write $\boldsymbol{z} = \widetilde{\boldsymbol{z}} + \boldsymbol{u}$.

For some intuition, we may wish to think of the special case of $\widetilde{\boldsymbol{z}}$ being the linear relation $\widetilde{\boldsymbol{z}} = \boldsymbol{X}\boldsymbol{\beta}$, where $\boldsymbol{X}$ can contain lagged dependent variables, so that we cover autoregressive distributed-lag models (e.g. used in co-integration analysis) as one of the special cases of $\boldsymbol{z} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$. This case focuses attention on the nonlinearity in the dynamics of the residuals $\boldsymbol{u}$, rather than on the functional form relating $\widetilde{\boldsymbol{z}}$ to $\boldsymbol{X}$. Naturally, our procedure will estimate simultaneously both $\boldsymbol{\beta}$ and the parameters governing the process $\boldsymbol{u}$.

10

We have touched on long memory in a footnote in the introduction. Here we give a short description of it. More details are given in Beran (1992), Robinson (1994), Baillie (1996), Abadir and Taylor (1999). The ACF $\rho_1, \rho_2, \ldots$ of a process $\{u_t\}_{t=1}^T$ is the sequence of correlations of the variable with its τ-th lag:

$$\rho_\tau \equiv \frac{\text{cov}\,(u_t, u_{t-\tau})}{\sqrt{\text{var}\,(u_t)\,\text{var}\,(u_{t-\tau})}}, \tag{7}$$

where $\rho_0 \equiv 1$. Long memory is a case where this function of τ decays very slowly as τ increases, typically hyperbolically and hence much slower than the exponential rate of decay obtained for stationary AR models. Unlike unit root models, shocks to a long-memory process do not have an everlasting impact.

The autocorrelation matrix of $\boldsymbol{u}$ can be written as

$$\boldsymbol{R} \equiv \begin{pmatrix} 1 & \rho_1 & \rho_2 & \ddots & \rho_{T-2} & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \ddots & \ddots & \rho_{T-2} \\ \rho_2 & \rho_1 & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \rho_1 & \rho_2 \\ \rho_{T-2} & \ddots & \ddots & \rho_1 & 1 & \rho_1 \\ \rho_{T-1} & \rho_{T-2} & \ddots & \rho_2 & \rho_1 & 1 \end{pmatrix}. \tag{8}$$

In the case of a simple AR(1) with autoregressive parameter ρ, we would have had $\rho_\tau = \rho^\tau$ and knowledge of ρ alone would have allowed filling the whole $\boldsymbol{R}$ matrix. In the general case, estimating $\boldsymbol{R}$ requires estimating $\rho_1, \ldots, \rho_{T-1}$, that is $T-1$ parameters. However, we have only T data points. Our solution is to let ρ_τ take the functional form in Abadir, Caggiano, and Talmain (2006)

$$\rho_\tau \approx \frac{1 - a\,[1 - \cos(\omega\tau)]}{1 + b\tau^c}, \tag{9}$$

with only 4 parameters to estimate rather than $T-1$. We find this to be a very accurate and parsimonious description of the ACFs in Figures 1 and 2, where we superimpose the empirical ACF and the one fitted to the original

data by means of (9).[4] The context in Abadir et al. (2006) is that of a single variable in the ACF domain, where they show that this 4-parameter functional form represents the dynamics of individual macroeconomic variables more accurately than $\mathrm{AR}(p)$ models. Here, we introduce the idea of incorporating ACFs into multivariate time-domain estimation in order to extract the relation linking the variables together while avoiding spurious correlations.

We now present a quasi Maximum Likelihood (ML) procedure to estimate jointly the parameters of $\widetilde{z}_t$ and the ACF of u_t. The adjective "quasi" is standard and indicates that the likelihood function is based on the asymptotic normality that arises from the central limit theorem. This covers a wide range of distributions, just requiring the first two moments of $\boldsymbol{u}$ to be finite. To simplify the exposition, we adopt the linear model

$$\boldsymbol{z} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \text{ with } \boldsymbol{u} \sim \mathrm{D}\left(\boldsymbol{0}, \boldsymbol{\Sigma}\right), \tag{10}$$

where $\boldsymbol{\Sigma}$ is the $T \times T$ autocovariance matrix. We will show how to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ jointly. We need to start by explaining a couple of features of the model, which we do in the next two paragraphs.

There are two implications to u_t being mean-reverting. First, $\boldsymbol{\Sigma}$ is proportional to the autocorrelation matrix $\boldsymbol{R}$ in (8)–(9), with $b, c > 0$ in (9). Second, the ML estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are consistent: the estimates converge to their true value as the sample size increases.[5]

We will assume that the sample mean of each variable in (10) has been subtracted for numerical convenience, for a reason that is explained in Remark 4 of the Appendix. We also assume that $\boldsymbol{X}$ is weakly exogenous (see Engle, Hendry, and Richard, 1983) for the parameters of (10), which is justified in our two applications but need not be so in general. Otherwise, one needs joint modelling of $\boldsymbol{z}$ and $\boldsymbol{X}$, or estimation of the parameters of (10) by means of instruments (effectively an orthogonal decomposition of the equation). Note that estimating parameters and testing hypotheses requires

12

weak exogeneity, but forecasting requires strong exogeneity; see Engle et al. (1983, p.286).

Now, to the joint estimation of all the parameters of (10). For any given $\boldsymbol{R}$, define

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} \equiv \left(\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{z} \tag{11}$$

as a function of $\boldsymbol{R}$. Denoting the determinant of a matrix $\boldsymbol{M}$ by $|\boldsymbol{M}|$, the Appendix shows that the ML Estimator (MLE) of $\boldsymbol{R}$ is obtained by maximizing

$$-\log\left|\left(\boldsymbol{z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}\right)'\boldsymbol{R}^{-1}\left(\boldsymbol{z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}\right)\boldsymbol{R}\right| \tag{12}$$

with respect to the parameters of the ACF: the optimization of the joint likelihood (for $\boldsymbol{\Sigma}$ and $\boldsymbol{\beta}$) now depends on only 4 parameters that are given in (9) and that determine the whole autocorrelation matrix $\boldsymbol{R}$. Once the optimal value $\widehat{\boldsymbol{R}}$ of $\boldsymbol{R}$ is obtained, the MLE of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} \equiv \widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{R}}}$.

The main novelty in our procedure of this section is the parsimonious parameterization of the autocorrelation matrix $\boldsymbol{R}$. This transcends the chosen estimation method, and it can be used as an input for methods other than ML. We chose ML because of its statistical optimality properties, but other choices are feasible and can employ our parameterization of $\boldsymbol{R}$. One such additional method is given in the Appendix.

III. Uncovering the Uncovered Interest Parity Regression

Using our procedure on the US\$-UK£ data set, we get the joint estimates for the adjustment dynamics of u_t as

$$\widehat{\rho}_\tau = \frac{1 - 1.07\left[1 - \cos\left(0.054\tau\right)\right]}{1 + 0.022\tau^{0.95}} \tag{13}$$

and the relation of interest as

$$\begin{aligned}
\widehat{s}_{t+1} \quad &= \quad -0.41 f_t \quad +1.11\, s_t + \widehat{u}_{t+1}. \\
&\qquad (-0.69) \quad (0.55) \\
&\qquad [-0.58] \quad [0.45]
\end{aligned} \tag{14}$$

13

Again, we do not report the constant. We can see that the HAC adjustment now makes almost no difference to the t-ratios for the tests that the parameters of f and s are 1 and 0, respectively, so there is very little residual dynamics or heteroskedasticity left over. Both estimated coefficients are not significantly different from their hypothesized values, as theory had predicted. The procedure has changed the story in a dramatic way. Once the dynamics of the problem are properly taken care of, neither the forward premium nor its components have any predictive power for the excess returns r_{t+1}. We cannot reject the hypothesis that $r_{t+1} = u_{t+1}$. Notice that the power of τ in (13) indicates that the memory in u_t decays eventually (unlike in unit-root models), but it does so more slowly than stationary linear AR models can allow. It must be stressed that we have shown that r_{t+1} is not related to the forward premium, but we have given no explanation here for the fact that r_{t+1} has such persistent dynamics of its own.

The contrast with the earlier regression is even more striking if we compare the bivariate scatter plots of r_{t+1} and $(f_t - s_t)$, before and after taking into account the ACF, in Figures 3 and 4, respectively. With the original data, there seemed to be a clear negative relation between r_{t+1} and $(f_t - s_t)$. After the ACF-implied transformation,[6] it looks like a round 'cloud' with no particular tendency or deformation, which is the way that the scatter plot of two independent variables should be, after standardization of the scale for both axes. The relation found at the beginning between these variables was spurious. It was just an artifact of the long memory of the variables involved in the regression, and the incomplete modelling of these dynamics. Of course, we could improve our results further by including risk premia, transaction costs, and/or peso problems. What we have done here is to show that our simple but unconventional dynamics already provide a lot of explanatory power for the series.

Our results do not contradict the possibility of carry trades that are

profitable, on average. They identify the source of profitability and increase the precision of the predictions by improving the fit of the equation. Our regression reveals that the source of excess returns is the momentum of the exchange rate rather than the nonzero forward premium. Hence, even when $f_t \approx s_t$ and there is no opportunity for a carry trade, there are profit opportunities due to the momentum of the exchange rate itself. The momentum happens to be coincidental with a nonzero forward premium, hence the earlier result when the dynamics were omitted from the regression and the forward premium was acting as a proxy for them.

IV. The Stock Market Application

In a monopolistically competitive economy, firms will make an economic profit, which can be thought of as dividends in a Modigliani-Miller setting. In equilibrium, the value of a firm is equal to the (stochastically) discounted flow of its dividends. However, these dividends are not completely exogenous in a general equilibrium framework. Talmain (2006) showed that, on a balanced growth path, the capitalization of the stock market should be proportional to GDP.

In the second application of our estimation procedure, we show that the S&P 500 index does not have the unit root often debated in the literature on stock market efficiency. Rather, it grows in persistent long cycles guided by the trend line given by GDP, as seen in Figure 5 where both variables are in logarithms of real US$. We gave references in the introduction for papers documenting the short-run momentum followed by a longer-run correction, as well as the connection with the underlying state of the economy. In addition, Cavaliere (2001) devised a nonparametric test (but no econometric model) that detected the presence of long memory in S&P 500 and rejected the unit root hypothesis. Also, Hong and Lee (2005) have devised tests which

15

they applied, demonstrating significant nonlinearities in the evolution of the mean of returns on S&P 500 and NASDAQ.

In this section, we study this relation by two different methods: the standard one (unit roots and co-integration analysis) and our procedure. We also show that standard methods would end up with estimated parameters that are unstable, often oscillating as more data becomes available, trying to mimic the nonlinearity of the process. S&P turns out to be quite different from linear integrated processes, hence co-integration cannot apply and different tools are needed to study this variable and its relation to the state of the economy.

A. The Estimates from Standard Analysis

Define y_t and s_t to be the logarithms of the annual real GDP and of the real stock market index, respectively.[7] Consider the error-correction model of s_t in terms of y_t,

$$
\begin{aligned}
\Delta s_t \;=\;& \alpha + \left(\beta_1 \Delta s_{t-1} + \cdots + \beta_p \Delta s_{t-p}\right) & (15) \\
& + \left(\gamma_1 \Delta y_{t-1} + \cdots + \gamma_q \Delta y_{t-q}\right) - \delta\left(s_{t-1} - y_{t-1}\right) + \delta_1 y_{t-1} + u_t,
\end{aligned}
$$

where the are no contemporaneous variables on the right-hand side, apart from u_t. If $\delta \neq 0$, the Error-Correction Mechanism (ECM) is defined by the expression

$$
-\delta\left(s_{t-1} - y_{t-1}\right) + \delta_1 y_{t-1} = -\delta\left(s_{t-1} - \lambda y_{t-1}\right), \qquad (16)
$$

where $\lambda \equiv 1 + \delta_1/\delta$. The ECM represents the long-run relation between s and y in 'equilibrium': $s_t^{\mathrm{e}} = \kappa + \lambda y_t^{\mathrm{e}}$, where κ is some constant. The long-run proportionality between s_t^{e} and y_t^{e} can be investigated by testing the hypothesis H_0: $\delta_1 = 0$. Let d_{t-1} be the deviation at time $t-1$ of s_{t-1} from its equilibrium value s_{t-1}^{e}, that is, $d_{t-1} \equiv s_{t-1} - s_{t-1}^{\mathrm{e}}$. The model implies that such a deviation will pull s_t back towards its long-term equilibrium

value by δd_{t-1} when $\delta > 0$. A small δ indicates a weak tendency for return to the long-term equilibrium. For more details, see Hendry, Pagan, and Sargan (1984).

Hendry and Von Ungern-Sternberg (1980) generalize this model to include an Integral Correction Mechanism (ICM), where cumulative imbalances of $(s_{t-1} - \kappa - y_{t-1})$ play a role in the catching-up of s_t with its equilibrium value. This is an important mechanism. When equity prices increase faster than their fundamental value, they build up the wealth of the stockholders faster than on a balanced growth path. This wealth buildup creates the condition for a tendency-reversal when stockholders convert their capital gains into additional discretionary spendings. Other factors, such as capital accumulation, pull in the same direction. We summarize all of these factors in one variable, the adjustment overhang, which is the cumulative sum of departures of the log of S&P from its fundamentals. Define

$$x_{t-1} \equiv \sum_{j=1}^{t-1} (s_j - y_j). \tag{17}$$

Adding this explanatory variable to the right-hand side of (15), we also need to add a balancing linear trend, in case the relation between s_j and y_j contains an intercept κ, because $\sum_{j=1}^{t-1} (s_j - \kappa - y_j) = \kappa (t-1) + x_{t-1}$. An augmented Dickey-Fuller t-test of -2.61 rejects the hypothesis that x_t has a unit root.

For S&P 500 over 1953–2004, we obtained the regression

$$
\begin{aligned}
\widehat{\Delta s_t} \;=\; & 0.0051\ \Delta s_{t-1} && -0.39\ \Delta s_{t-2} && -0.33\ \Delta s_{t-3} \\
& (0.04) && (-2.67) && (-2.21) \\[4pt]
& -0.046\ x_{t-1} && -0.0029\ t && +1.62\ \Delta y_{t-2}, \\
& (-4.73) && (-2.30) && (1.81)
\end{aligned}
\tag{18}
$$

where the t-ratios for significance from 0 are in parentheses. Insignificant variables have been dropped (except Δs_{t-1} which we will discuss in the next paragraph), and we do not report the constant.

The regression indicates that the proportionality hypothesis holds, and that the ICM plays a more important role than the ECM. The fit is $R^2 = 37.4\%$. Closer inspection reveals the fragility of the estimates. In Figure 6, each central line presents recursive parameter estimates as the sample is increased to its full size, with ± 2 standard-error bands for approximate 95% confidence intervals. Panels 1–6 represent the parameters of the variables in the order that they appear in the regression. For stability of the parameter estimates, the central lines should be nearly horizontal. Significance of these estimates occurs when the bands do not include zero. The bands for the initial estimates are understandably large, since these are based on the very few first observations. Otherwise, we make the following observations.

Although all but one reported final parameter estimate are significant, this is not systematically so throughout the period. A sample stopping a few years short would have found the first estimate significant, with the third and fifth insignificant. The sixth is marginal throughout, at the 95% significance level. Panels 1, 3, 4, and 5 show parameters that are breaking up then down and so on, trying to mimic the nonlinearity of the time path of S&P seen earlier in Figure 5. This series of 'breaks'[8] in alternating directions is predictable and can be avoided as we shall see.

Of course, one can include more lags of the variables in question, but this worsens rather than improves stability, while not improving the fit since these additional estimates are insignificant. Also, the RESET test is 2.69, with a borderline p-value of 10.8%, indicating potentially some omitted non-linearities.

B. The Estimates from our ACF-Based ML Procedure

Once we consider the ACF of s_t, it becomes clear that s_t is not a unit-root process, but rather the nonlinear long-memory type discussed earlier. Fitting to it our theoretical functional form (9), we get the curve in Figure

2 where the approximation is again very good.

For the same sample, we get the joint estimates for the adjustment dynamics of u_t as

$$\widehat{\rho}_\tau = \frac{1 - 1.04\,[1 - \cos(0.61\tau)]}{1 + 0.17\tau^{1.11}} \tag{19}$$

and the relation

$$
\begin{aligned}
\widehat{\Delta s_t} \;=\;\; & -0.71\;\Delta s_{t-1} && -0.60\;\Delta s_{t-2} && -0.53\;\Delta s_{t-3} \\
& (-5.57) && (-4.27) && (-4.58) \\[4pt]
& -0.069\;x_{t-1} && -0.0045\;t && -1.45\;\Delta y_{t-1} + \widehat{u}_t. \\
& (-5.69) && (-3.18) && (-2.38)
\end{aligned}
\tag{20}
$$

Insignificant variables have been dropped and the proportionality hypothesis is accepted as before. The fit for the returns on the stock market, Δs_t, is $R^2 = 50.8\%$, surprisingly large given that we have only used our dynamics and GDP as explanatory variables. The p-value for the improvement in fit compared to the standard model is 3.91%, a substantial improvement. The practical significance of this R^2 statistic is elucidated in Campbell and Thompson (2008). What is the gain to a mean-variance investor from identifying a predictor of the S&P 500 with this explanatory power? Let SR (our notation) be the usual Sharpe ratio. Their answer is that this gain depends on the risk aversion of this investor and, as an approximation, on the ratio $R^2/(\text{SR})^2$. Since the annual Sharpe ratio on the S&P 500 for the period 1953-2004 is 0.226, we can see that the ratio is

$$\frac{R^2}{(\text{SR})^2} = \frac{0.508}{0.0512} = 9.92. \tag{21}$$

Hence, the gains are considerable.[9]

Going back to our equation, the RESET is now 0.0004, with a p-value of 98.4%, clearly indicating no leftover nonlinearities. In contrast to the unit roots and co-integration analysis, the coefficients are now very stable throughout the sample and do not oscillate as before, and this is verified in Figure 7. Gone is the sequence of breaks up and down previously observed

for a parameter's estimate. The ACF has taken care of the long cycle that caused these undesirable features in the estimates of the traditional method. The previously omitted nonlinear dynamics of S&P 500, away from the long-run proportionality to GDP, have been accounted for by the ACF dynamics incorporated in u_t.

Apart from the memory pattern in (19), the residuals have no leftover dynamics: the LM test of no autocorrelation is highly insignificant, with a p-value of 82.5%. The power of τ in (19) implies that the adjustment dynamics in u_t are mean-reverting, but slower than stationary linear AR models. This means that s_t does not stray too far from the trend-line that y_t provides, and that it has the tendency to revert to it over time. In the context of Figure 5, GDP provides the stochastic trend line around which S&P 500 moves in long-memory cycles. Notice that these cycles have a pattern that is fairly well behaved, in contrast with the unpredictability that a unit-root model would have implied. As time passes, the variance bounds of a random walk would diverge linearly away from its trend line, and there would be no tendency for trend-reversion (zero probability of this happening). However, in Figure 5, we see no such feature, neither for GDP nor for S&P. The analysis of this section confirms it for S&P. For GDP, see Abadir and Talmain (2002).

Figure 8 shows the implied predictions of the percentage changes in S&P 500. The predictions do not shadow (or lag behind) the actual values, contrary to what standard time series methods yield. The actual changes are generally well reflected in our fitted values. There are a few notable exceptions that can be attributed to two main reasons: temporary S&P overreactions that are unjustified by fundamentals and/or model predictions that materialize after a short delay. For example, our model provides evidence of overreaction at the aggregate level, supporting the one discussed for individual stocks in the references listed in the introduction. The excess volatility

20

of S&P in the early 70's is detected by our predictions pointing to changes that are less pronounced than the actual ones. The implication was that, whenever the market deviated from our prediction, it undertook a correction immediately the next period, bringing it nearer to the fundamental values from our equation. For example, in 1971–1972, S&P was growing while our model was predicting that it should go down, well ahead of the oil shock. That happened a couple of years later, with an extremely large negative change. The oil shock was not completely responsible for that correction, as our modelling of the fundamentals has revealed. A similar story can be told about the end of the 1990's and early 2000's. Our prediction was pointing the way down at the heydays of the dot-com bubble and, as it burst so dramatically, our method showed that the fall of the market has been overdone. A subsequent upward correction took place, as anticipated by our prediction. Finally, the rise in GDP productivity that took place in the early 90's led our model to predict an increase in S&P, which actually happened with a delay of a couple of years.

In both applications of our method, once the long cycles are taken into account, the residuals were found to contain no structural breaks. We did not need to add to the models any dummy variables for breaks, even in periods where the relation in terms of the raw data seemed at first sight to break, e.g. due to Britain's exit of the exchange-rate mechanism or to extended S&P bubbles or corrections. In the case of the stock market, the ICM plays the following two roles. First, it makes sure that if S&P strays from GDP too fast, it is brought back faster than usual. This can be seen in the last four years, which correspond to the bubble that burst, and where the cycle was shorter than the previous ones. Second, and as a result, the ICM ensures that the S&P cycles are of random length (or frequency), once the two dynamics arising from ICM and ACF are added up.

V. Concluding Comments

Integration and co-integration have had a huge impact on the analysis of macroeconomic and aggregate financial data. They were a good first step in establishing methods to deal with variables containing much persistence. Here, we present a more general econometric method of analysis that is justified by the economic model of Abadir and Talmain (2002), and we show how it explains the evolution of exchange rates and S&P 500.

Our method has the potential to reveal new insights when two conditions are satisfied: whenever a model requires us to disentangle the dynamics of a dependent variable from its relation with other variables of interest, and when these dynamics are of the long-memory form. The first condition is the norm. The second feature is increasingly encountered, given the recent econometric advances in handling long memory processes and the evidence it has uncovered so far. For example, see the arguments in Abadir and Taylor (1999) and the numerical results in Diebold and Rudebusch (1989), Baillie and Bollerslev (1994, 2000), Gil-Alaña and Robinson (1997), Chambers (1998), Abadir and Talmain (2002), Abadir et al. (2006).

Appendix

We can use the Cholesky decomposition to write the matrix $\boldsymbol{\Sigma}$ of (10) as $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{L} \boldsymbol{L}'$, where $\boldsymbol{L}^{-1}$ is the lower triangular matrix that removes autocorrelation from $\boldsymbol{u}$ and takes the form

$$\boldsymbol{L}^{-1} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ -\boldsymbol{\alpha}' & 1 \end{pmatrix}, \tag{A1}$$

with $\boldsymbol{\alpha}' \equiv (\alpha_{T-1}, \cdots, \alpha_2, \alpha_1)$ and $\boldsymbol{A}$ a lower-triangular block of dimension $T-1$. Therefore, premultiplying (10) by $\boldsymbol{L}^{-1}$,

$$\boldsymbol{L}^{-1} \boldsymbol{z} = \boldsymbol{L}^{-1} \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{\varepsilon} \sim \mathrm{D} \left(\boldsymbol{0}, \ \sigma^2 \boldsymbol{I}_T \right). \tag{A2}$$

(Notice that we reserve ε_t for well-behaved errors, and use u_t for errors with possible patterns such as autocorrelation.) The transformed residuals $\boldsymbol{\varepsilon}$ are now uncorrelated, and standard estimation procedures can be applied to the transformed model.[10] The exogeneity assumption given before (11) implies that, given the past, ε_t is independent of the t-th row of $\boldsymbol{L}^{-1}\boldsymbol{X}$, since $\boldsymbol{L}^{-1}$ is lower triangular (see Remark 2 below).

The transformed data $\boldsymbol{L}^{-1}\boldsymbol{z}$ and $\boldsymbol{L}^{-1}\boldsymbol{X}$ can be regressed by traditional methods. The Cholesky decomposition command is built-in as standard in all matrix-handling languages, such as Gauss and Matlab. The Generalized Least Squares (GLS) estimators can be obtained by minimizing the criterion $(\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta})$ with respect to all parameters jointly. Alternatively, the ML estimators are obtained by maximizing

$$-\log|\boldsymbol{\Sigma}| - (\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta}), \tag{A3}$$

where only the first term differs from the GLS criterion, and it has the beneficial effect of ensuring that the elements of the diagonal of $\boldsymbol{L}^{-1}$ are not too far from unity. This difference is responsible for another desirable property that the method of ML has, that it is invariant to reparameterizations of the model. Concentrating the log-likelihood with respect to

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} \equiv (\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{z} \text{ and } \widehat{\sigma}_{\boldsymbol{L}}^2 \equiv \frac{1}{T}(\boldsymbol{z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})' \left(\boldsymbol{L}\boldsymbol{L}'\right)^{-1} (\boldsymbol{z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})$$

$$\tag{A4}$$

(i.e. substituting $\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}, \widehat{\sigma}_{\boldsymbol{L}}^2$ for $\boldsymbol{\beta}, \sigma^2$ into (A3) and using $\boldsymbol{\Sigma} \equiv \sigma^2 \boldsymbol{L}\boldsymbol{L}' \propto \boldsymbol{R}$) gives

$$-\log\left|\frac{1}{T}(\boldsymbol{z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})' \left(\boldsymbol{L}\boldsymbol{L}'\right)^{-1} (\boldsymbol{z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})\boldsymbol{L}\boldsymbol{L}'\right| - T \tag{A5}$$

$$= -\log\left|\frac{1}{T}(\boldsymbol{z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})'\boldsymbol{R}^{-1}(\boldsymbol{z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})\boldsymbol{R}\right| - T$$

$$= -\log\left|(\boldsymbol{z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})'\boldsymbol{R}^{-1}(\boldsymbol{z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})\boldsymbol{R}\right| + T\log(T) - T$$

by $\left|T^{-1}\boldsymbol{M}\right| = T^{-T}\left|\boldsymbol{M}\right|$ for any $T \times T$ matrix $\boldsymbol{M}$. Dropping the constant term $-T\left(1 - \log(T)\right)$ yields (12), to be optimized with respect to the para-

23

meters of the ACF. The log-likelihood is nonlinear in $\boldsymbol{R}$, and a grid search over the 4 parameters of the ACF may be needed to ensure that a global maximum is achieved.

Before we comment on the details of our procedure, we indicate how it grew out of the traditional treatment of models with autocorrelated errors, which are nested within our model. We take the simplest example

$$z_t \;=\; \gamma x_t + u_t, \tag{A6}$$

$$\text{with } u_t \;=\; \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1, \quad \varepsilon_t \sim \text{IID}\left(0, \sigma^2\right).$$

To estimate (A6), taking into account the autocorrelation of u_t, the variables of the first equation (z_t and x_t) are transformed, then they are regressed by OLS to estimate the parameter γ of the relation. The vector $\boldsymbol{z} \equiv (z_1, \ldots, z_T)'$ is transformed into

$$\boldsymbol{L}^{-1}\boldsymbol{z} \equiv \begin{pmatrix} \varphi & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\rho & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_T \end{pmatrix} = \begin{pmatrix} \varphi z_1 \\ z_2 - \rho z_1 \\ z_3 - \rho z_2 \\ \vdots \\ z_T - \rho z_{T-1} \end{pmatrix}, \tag{A7}$$

where an estimate of ρ is plugged in, and where φ is usually chosen as $\sqrt{1-\rho^2}$ to stabilize the variance of the transformed residuals. The lower triangular matrix $\boldsymbol{L}^{-1}$ that premultiplies the vector of z_t-values arises from the Cholesky decomposition

$$\boldsymbol{R} \equiv \begin{pmatrix} 1 & \rho & \rho^2 & \\ \rho & 1 & \rho & \ddots \\ \rho^2 & \rho & 1 & \ddots \\ & \ddots & \ddots & \ddots \end{pmatrix} = \left(1 - \rho^2\right) \boldsymbol{L}\boldsymbol{L}'. \tag{A8}$$

Together with $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{L}\boldsymbol{L}'$, we see that the proportionality factor linking $\boldsymbol{R}$ to $\boldsymbol{\Sigma}$ is $\sigma^2 / \left(1 - \rho^2\right)$, the asymptotic variance of u_t.

If u_t were following an AR(p), then the lower triangular matrix $\boldsymbol{L}^{-1}$ in (A7) would contain $p+1$ nonzero diagonals, and the first p rows would have a normalization as was done for φ; e.g. see Chapter 5 of Amemiya (1985). When the variables have long memory, as is in our case, one needs a very large p to make this transformation. We overcame this problem by using our new ACF-based method in a parsimonious way. Using the matrix companion form, Abadir, Hadri, and Tzavalis (1999) showed that long lags have a similar effect to adding dimensions to a VAR (Vector AR), which would increases the bias and variance of the estimators. Finding a parsimonious solution avoids these types of problems.

We make the following remarks on the requirements and/or features of $\boldsymbol{R}$ and the corresponding $\boldsymbol{L}$ in our procedure:

1. In estimating the parameters of the ACF, one needs to restrict their values so that the estimated $\widehat{\boldsymbol{R}}$ is positive definite, since this is true (by definition) for $\boldsymbol{R}$. There is no explicit formula for this restriction, because there is no explicit solution for the roots of polynomials of order greater than 4. Nevertheless, it is straightforward to implement the restriction numerically either by skipping solutions that do not satisfy the restriction, or by imposing a large Lagrangian penalty in the objective (e.g. log-likelihood) function.

2. The lower triangularity of $\boldsymbol{L}^{-1}$ ensures that each element of the transformed $\boldsymbol{z}$ is constructed only from past and current (but no future) values of z_t; e.g. see (A7). The same comment applies to $\boldsymbol{X}$.

3. The elements in the last row of $\boldsymbol{L}^{-1}$ have an interpretation as the coefficients of an AR($T-1$) representation for the last transformed data point, which is why we stated them explicitly in (A1). One may wish to restrict the optimization procedure such that it produces stationary roots for this AR representation, but we have not done so. Note that

any non-explosive process, whether nonlinear and/or nonstationary, can be represented as an invertible MA having time-varying coefficients, which explains the time-varying AR representations implied by the rows of $\boldsymbol{L}^{-1}$. This is known in time series as Cramér's decomposition, a generalization of Wold's decomposition, and it explains how the nonlinear process of Abadir and Talmain (2002) can be estimated by our linear representation procedure.

4. A well-known feature of the transformed model (A2) is that the constant, once transformed by any $\boldsymbol{L}^{-1}$, is not a constant vector anymore; e.g. use $\boldsymbol{\iota} \equiv (1, \ldots, 1)'$ instead of $\boldsymbol{z}$ in (A7) and compare the first element to the remaining $T - 1$. In our procedure, it is therefore assumed that the data ($\boldsymbol{z}$ and $\boldsymbol{X}$) have been de-meaned before being transformed. This is because the procedure is based on transforming vectors, say $\boldsymbol{y}$, centered around $\boldsymbol{0}$ from $\boldsymbol{y} \sim \mathrm{D}(\boldsymbol{0}, \boldsymbol{L}\boldsymbol{L}')$ into $\boldsymbol{L}^{-1}\boldsymbol{y} \sim \mathrm{D}(\boldsymbol{0}, \boldsymbol{I}_T)$. Having a nonzero sample mean in $\boldsymbol{y}$ would have introduced a common factor of $\boldsymbol{L}^{-1}\boldsymbol{\iota}$ in all these transformed variables, which may dominate these series and produce some seemingly common factor that causes multicollinearity and other unnecessary numerical instabilities. If a constant is required in the regression, it should be transformed separately then added to the regression for transformed variables. Numerical instabilities apart, the theorem of Frisch and Waugh (1933) proves that the resulting point estimates would be identical with or without removing the mean.

REFERENCES

Abadir, Karim M., Giovanni Caggiano, and Gabriel Talmain, 2006, Nelson-Plosser revisited: the ACF approach, working paper, University of Glasgow (http://www.gla.ac.uk/media/media_22198_en.pdf).

Abadir, Karim M., Kaddour Hadri, and Elias Tzavalis, 1999, The influence of VAR dimensions on estimator biases, *Econometrica* 67, 163-181.

Abadir, Karim M. and Jan R. Magnus, 2002, Notation in econometrics: a proposal for a standard, *Econometrics Journal* 5, 76-90.

Abadir, Karim M., and Gabriel Talmain, 2002, Aggregation, persistence and volatility in a macro model, *Review of Economic Studies* 69, 749-779.

Abadir, Karim M., and A.M. Robert Taylor, 1999, On the definitions of (co-)integration, *Journal of Time Series Analysis* 20, 129-137.

Aldrich, John, 1995, Correlations genuine and spurious in Pearson and Yule, *Statistical Science* 10, 364-376.

Amemiya, Takeshi, 1985, *Advanced Econometrics* (Harvard University Press, Cambridge).

Apte, Prakash, Piet Sercu, and Raman Uppal, 2004, The exchange rate and purchasing power parity: extending the theory and tests, *Journal of International Money and Finance* 23, 553-571.

Backus, David K., Allan W. Gregory, and Chris I. Telmer, 1993, Accounting for forward rates in markets for foreign currency, *Journal of Finance* 48, 1887-1908.

Baillie, Richard T., 1996, Long memory processes and fractional integration in econometrics, *Journal of Econometrics* 73, 5-60.

Baillie, Richard T., and Tim Bollerslev, 1994, Cointegration, fractional cointegration, and exchange rate dynamics, *Journal of Finance* 49, 737-745.

Baillie, Richard T., and Tim Bollerslev, 2000, The forward premium anomaly is not as bad as you think, *Journal of International Money and Finance* 19, 471-488.

Bekaert, Geert, 1996, The time variation of risk and return in foreign exchange markets: a general equilibrium perspective, *Review of Financial Studies* 9, 427-470.

Beran, Jan, 1992, Statistical methods for data with long-range dependence, *Statistical Science* 7, 404-427 (with discussion).

Bhojraj, Sanjeev and Bhaskaran Swaminathan, 2006, Macromomentum: returns predictability in international equity indices, *Journal of Business* 79, 429-451.

Campbell, John Y., and Robert J. Shiller, 1988, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661-676.

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: can anything beat the historical average? *Review of Financial Studies* 21, 1509-1531.

Cavaliere, Giuseppe, 2001, Testing the unit root hypothesis using generalized range statistics, *Econometrics Journal* 4, 70-88.

Chambers, Marcus J., 1998, Long memory and aggregation in macroeconomic time series, *International Economic Review* 39, 1053-1072.

Chordia, Tarun and Lakshmanan Shivakumar, 2006, Earnings and price momentum, *Journal of Financial Economics* 80, 627-656.

Cochrane, John H., 2008, The dog that did not bark: a defense of return predictability, *Review of Financial Studies* 21, forthcoming.

Cooper, Michael J., Roberto C. Gutierrez, and Allaudeen Hameed, 2004, Market states and momentum, *Journal of Finance* 59, 1345-1365.

De Bondt, Werner F.M., and Richard Thaler, 1985, Does the stock market overreact? *Journal of Finance* 40, 793-808.

De Bondt, Werner F.M., and Richard Thaler, 1987, Further evidence on investor overreaction and stock market seasonality, *Journal of Finance* 42, 557-581.

De Bondt, Werner F.M., and Richard Thaler, 1989, A mean-reverting walk down Wall Street, *Journal of Economic Perspectives* 3, 189-202.

Diebold, Francis X., Javier Gardeazabal, and Kamil Yilmaz, 1994, On cointegration and exchange rate dynamics, *Journal of Finance* 49, 727-735.

Diebold, Francis X., Steven Husted, and Mark Rush, 1991, Real exchange rates under the gold standard, *Journal of Political Economy* 99 1252-1271.

Diebold, Francis X., and Glenn D. Rudebusch, 1989, Long memory and persistence in aggregate output, *Journal of Monetary Economics* 24, 189-209.

Engel, Charles, 1996, The forward discount anomaly and the risk premium: a survey of recent evidence, *Journal of Empirical Finance* 3, 123-192.

Engel, Charles, and James D. Hamilton, 1990, Long swings in the dollar: are they in the data and do markets know it? *American Economic Review* 80, 689-713.

Engle, Robert F., David F. Hendry, and Jean-François Richard, 1983, Exogeneity, *Econometrica* 51, 277-304.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23-49.

Fisher, Irving, 1930, *The Theory of Interest* (Macmillan, New York).

Frisch, Ragnar, and Frederick V. Waugh, 1933, Partial time regressions as compared with individual trends, *Econometrica* 1, 387-401.

Froot, Kenneth A., and Richard H. Thaler, 1990, Foreign exchange, *Journal of Economic Perspectives* 4, 179-192.

Gil-Alaña, Luis A., and Peter M. Robinson, 1997, Testing of unit root and other nonstationary hypotheses in macroeconomic time series, *Journal of Econometrics* 80, 241-268.

Griffin, John M., Xiuqing Ji, and J. Spencer Martin, 2003, Momentum investing and business cycle risk: evidence from pole to pole, *Journal of Finance* 58, 2515-2547.

Grinblatt, Mark, and Tobias J. Moskowitz, 2004, Predicting stock price movements from past returns: the role of consistency and tax-loss selling, *Journal of Financial Economics* 71, 541-579.

Grundy, Bruce D., and J. Spencer Martin, 2001, Understanding the nature of the risks and the source of the rewards to momentum investing, *Review of Financial Studies* 14, 29-78.

Hendry, David F., Adrian R. Pagan, and J. Denis Sargan, 1984, Dynamic specification, in Z. Griliches and M.D. Intriligator, eds.: *Handbook of Econometrics*, vol. 2 (North-Holland, Amsterdam).

Hendry, David F., and Thomas Von Ungern-Sternberg, 1980, Liquidity and inflation effects on consumers' expenditure, in A. S. Deaton, ed.: *Essays in the Theory and Measurement of Consumers' Behaviour* (Cambridge University Press, Cambridge). Reprinted In Hendry, D. F., 1993, *Econometrics: Alchemy or Science?* (Basil Blackwell, Oxford).

Hong, Yongmiao, and Yoon Jin Lee, 2005, Generalized spectral tests for conditional mean models in time series with conditonal heteroscedasticity of unknown form, *Review of Economic Studies* 72, 499-541.

Hvidkjaer, Soeren, 2006, A trade-based analysis of momentum, *Review of Financial Studies* 19, 457-491.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: implications for stock market efficiency, *Journal of Finance* 48, 65-91.

Jegadeesh, Narasimhan, and Sheridan Titman, 2001, Profitability of momentum strategies: an evaluation of alternative explanations, *Journal of Finance* 56, 699-720.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541-1578.

Lee, Charles M.C., and Bhaskaran Swaminathan, 2000, Price momentum and trading volume, *Journal of Finance* 55, 2017-2069.

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* 56, 815-849.

Lewis, Karen K., 1995, Puzzles in international financial markets, in G. Grossman and K. Rogoff, eds.: *Handbook of International Economics*,

vol. 3 (North-Holland, Amsterdam).

Liew, Jimmy, and Maria Vassalou, 2000, Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics* 57, 221-245.

Okunev, John, and Derek White, 2003, Do momentum-based strategies still work in foreign currency markets? *Journal of Financial and Quantitative Analysis* 38, 425-447.

Petkova, Ralitsa, 2006, Do the Fama-French factors proxy for innovations in predictive variables? *Journal of Finance* 61, 581-612.

Robinson, Peter M., 1994, Time series with strong dependence, in C.A. Sims, ed.: *Advances in Econometrics: sixth world congress*, vol.1 (Cambridge University Press, Cambridge).

Sarno, Lucio, Giorgio Valente, and Hyginus Leon, 2006, Nonlinearity in deviations from uncovered interest parity: an explanation of the forward bias puzzle, *Review of Finance* 10, 443-482.

Talmain, Gabriel, 2006, Stock market valuation and monopolistic competition: a dynamic stochastic general equilibrium approach, working paper, University of Glasgow (http://www.gla.ac.uk/media/media_26604_en.pdf).

Vassalou, Maria, 2003, News related to future GDP growth as a risk factor in equity returns, *Journal of Financial Economics* 68, 47-73.

Notes

[1] For a historical account of the idea of spurious correlations, see Aldrich (1995). Clearly, the idea is not restricted to the setting of unit roots, and it predates this literature by far. Spurious correlations can occur even if the series have dynamic properties that are different from unit-root series.

[2] A time series is said to have long memory if its autocorrelations dampen very slowly, more so than the exponential decay rate of stationary autoregressive models but faster than the permanent memory of unit roots. Unlike the latter, long-memory series revert to their (possibly trending) means.

[3] A full modelling of S&P should look at the further explanatory power that other variables have, but this is not the purpose of this paper.

[4] In the illustrative Figures 1 and 2, the ACFs were fitted directly to the variables. In terms of (6), these figures set $\widetilde{z}_t = 0$, a restriction that will not be imposed in (13) and (19).

[5] The asymptotic distribution will depend on the properties of the variables z, X in general, but if the estimated residuals are found to satisfy $c > \frac{1}{2}$ (implying square-summability of ρ_τ), then standard tests are justified asymptotically. This condition on c is sufficient but not necessary, and allows us to skip the explicit specification of the time-series properties of z, X.

[6] This is the terminology used for the presence of R in (11). See the Appendix for a further explanation.

[7] We use the S&P 500 as our stock index. The other variables we need are an aggregate price index and real output which, unlike a stock index, is only available at low frequency. To avoid this seasonality problem interfering with the results, we use annual data. Real annual GDP is available from National Accounts. The consumer price index CPI is also available from the same source at a monthly frequency. Theoretically, the value of firms is related at each moment in time with current output. However the figure for annual

GDP is the production throughout the year. Hence, it must be related to an average stock index over the year. We first construct the average daily close of the S&P 500 at the highest frequency at which the CPI is available. Next, we divide this average by the CPI to obtain an index of the real value of the stock market. Finally, we convert this real stock market index into a yearly index.

[8]Part of the definition of a break is that it is an exogenous (and infrequent) event, and is an admission of failure of a model to explain a feature of the data.

[9]An investor, who has identified a predictor, can use this knowledge to reduce the (conditional) variance of the return on the S&P 500. Hence, her optimal behavior will be to invest more in the stock index and less in the risk-free asset, increasing the expected return of her optimal portfolio.

[10]Analysis of the estimates of the error term e will determine the model's adequacy. Such diagnostics include checking for leftover persistence that can be due to a spurious relation between z and X. For an illustration of these checks, see the two applications given earlier.
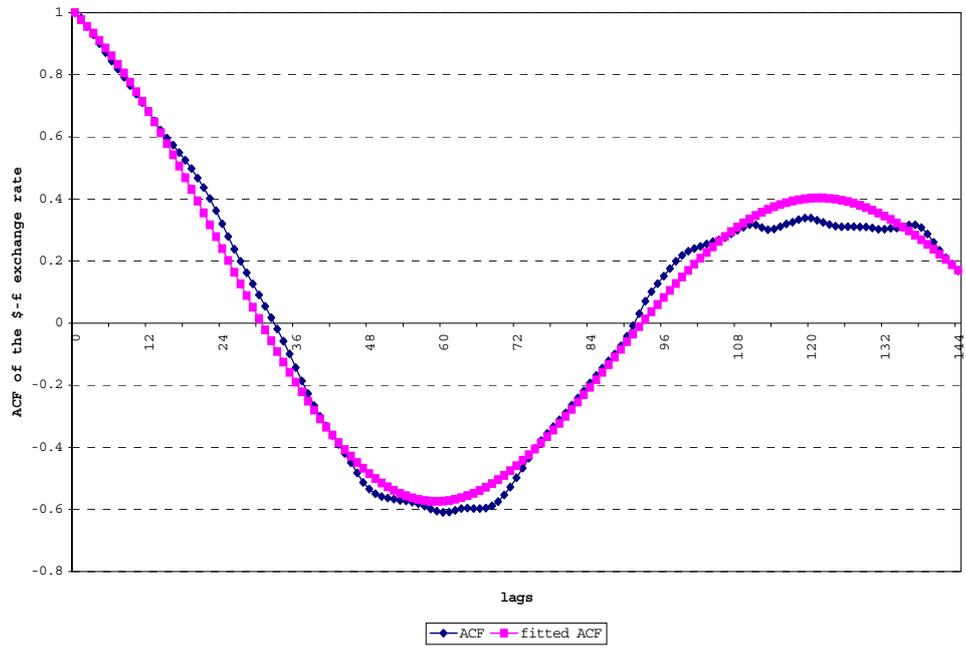
Figure 1. ACF of the logarithm of the $/£ exchange rate and its fit by our functional form.
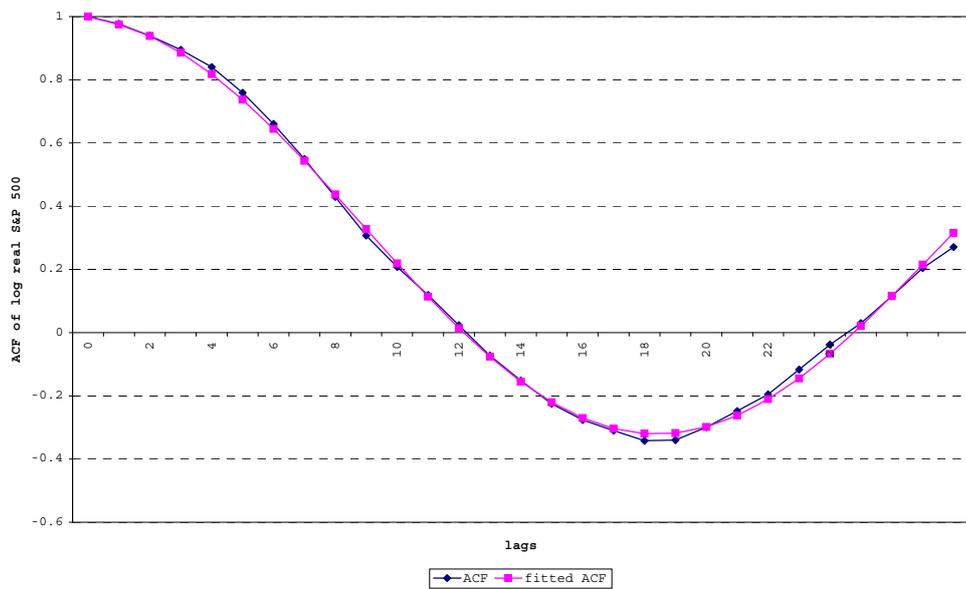

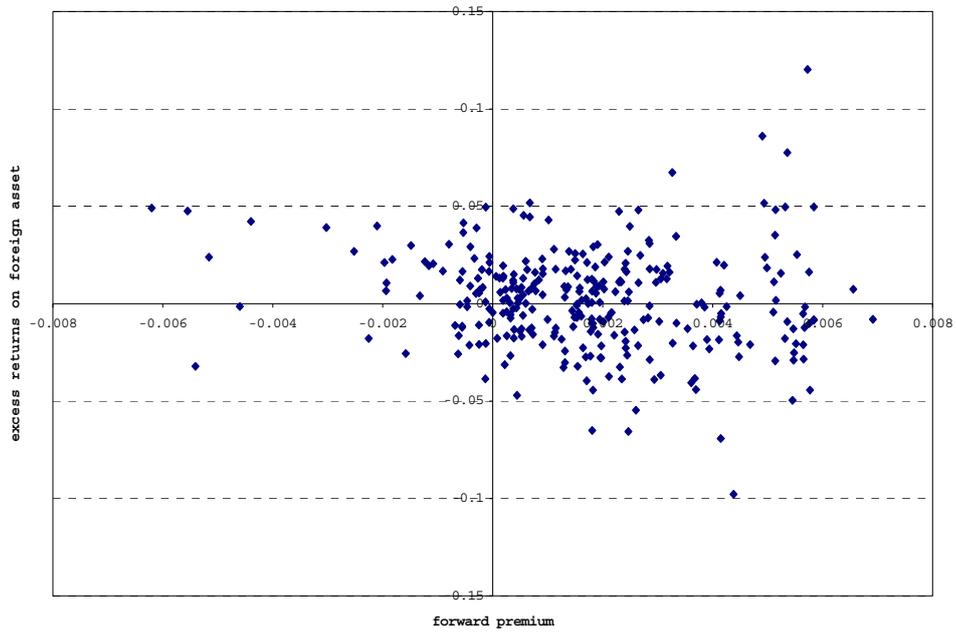**Figure 2**. ACF of the logarithm of S&P 500 and its fit by our functional form.

Figure 3. Scatter plot of the original data on excess returns r_{t+1} vs. the forward premium $(f_t - s_t)$, for equation (2).
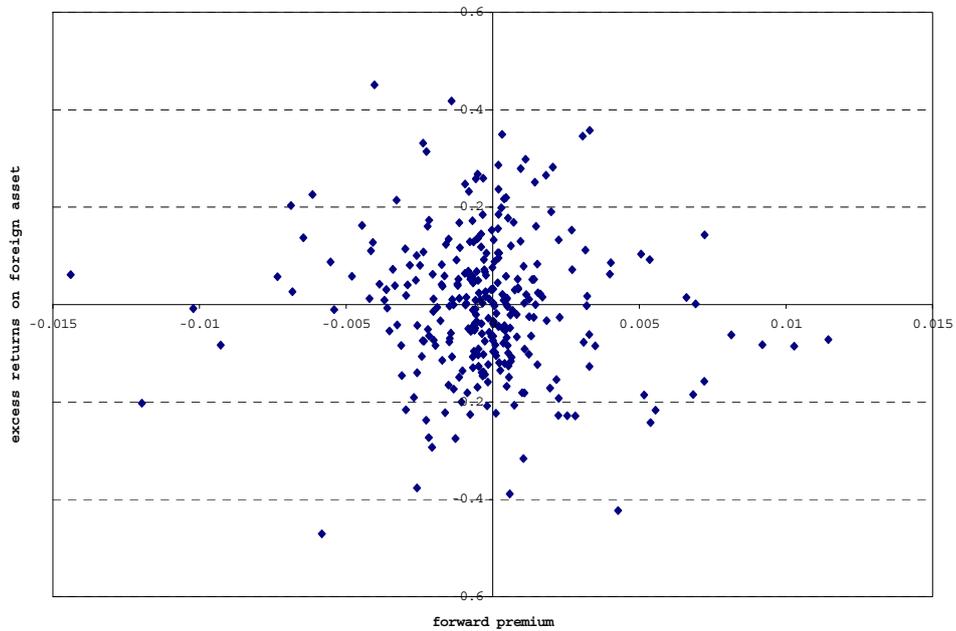

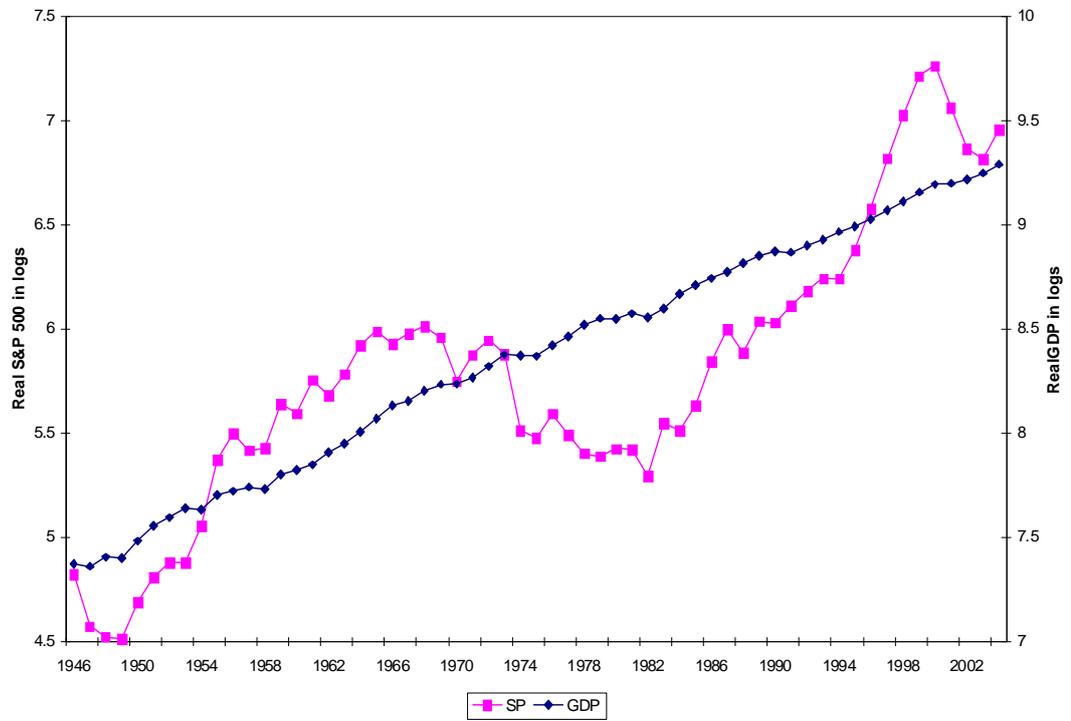**Figure 4**. Scatter plot of the ACF-transformed excess returns r_{t+1} vs. the forward premium $(f_t \overline{{}_{36}} s_t)$, for equation (2).
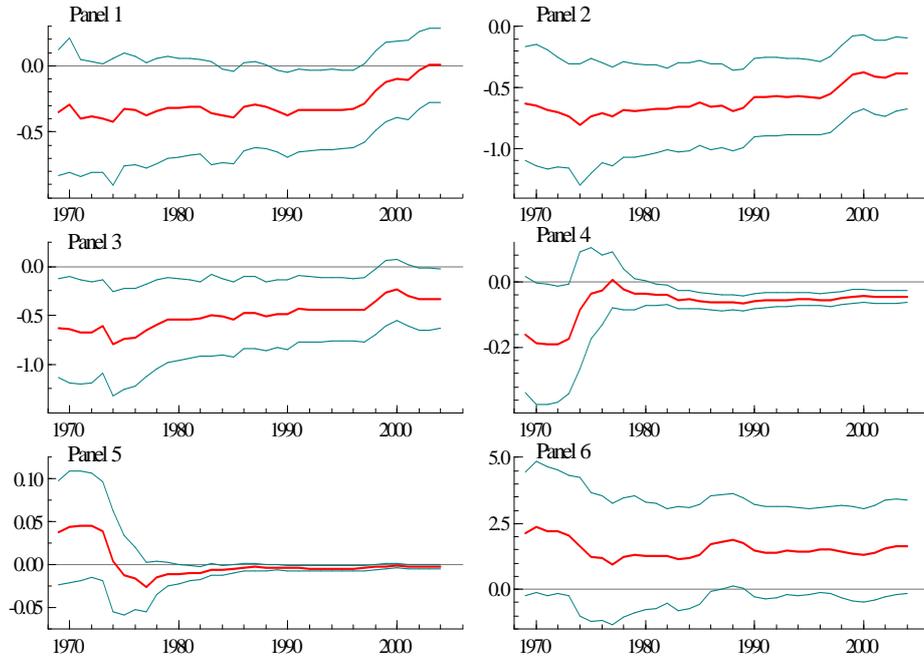
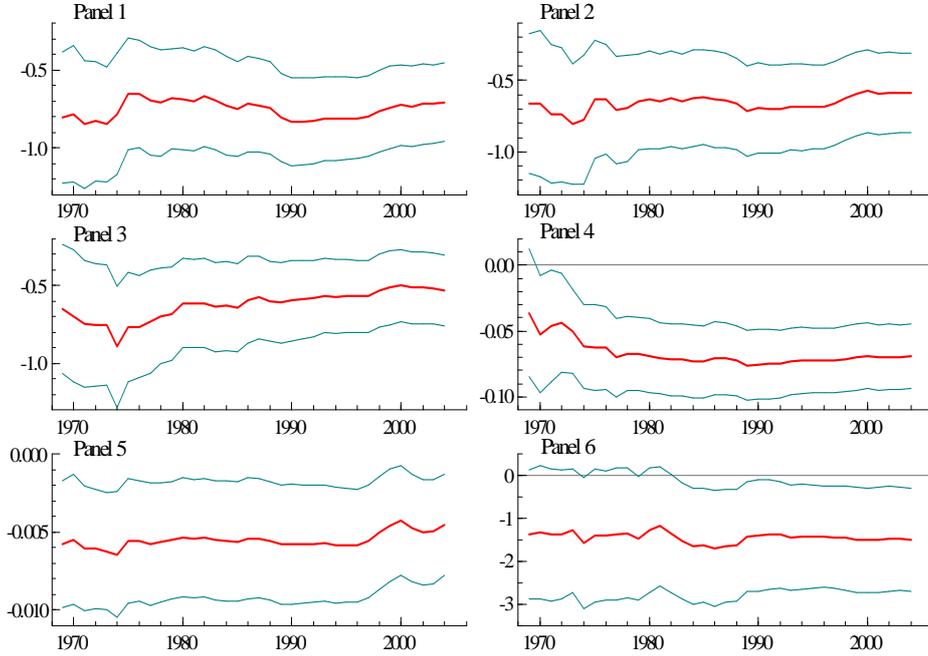**Figure 5**. Logarithms of real S&P 500 and real GDP.

Figure 6. Recursive parameter estimates as the sample size is increased, with ±2 standard-error bands. Panels 1–6 represent the parameter estimates of Δs_{t-1}, Δs_{t-2}, Δs_{t-3}, x_{t-1}, t, Δy_{t-2}, respectively.

Figure 7. Recursive parameter estimates as the sample size is increased, with ±2 standard-error bands. Panels 1–6 represent the parameter estimates of Δs_{t-1}, Δs_{t-2}, Δs_{t-3}, x_{t-1}, t, Δy_{t-1}, respectively.
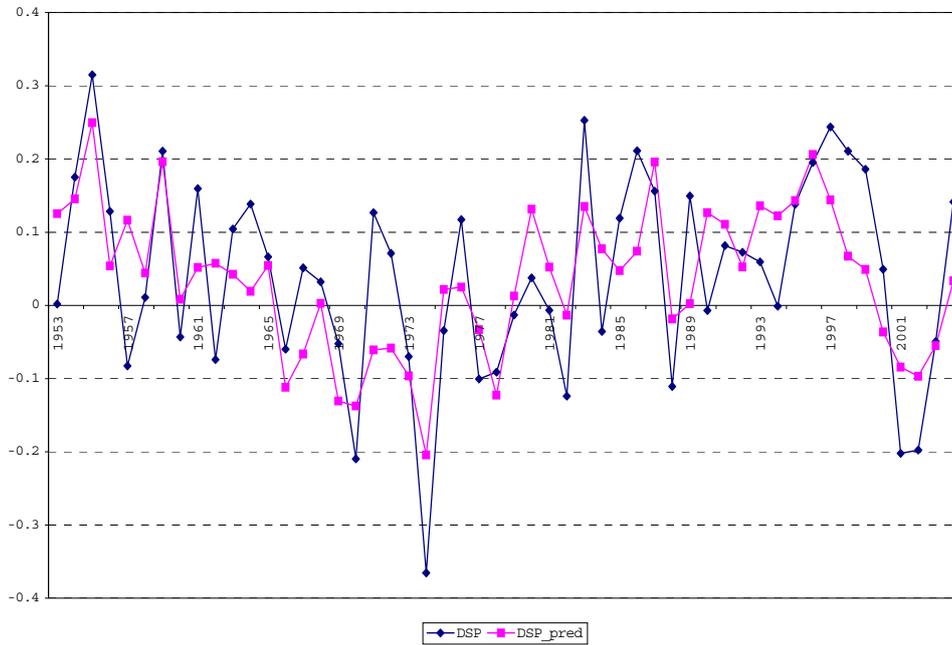
Figure 8. Actual and predicted changes in the logarithm of real S&P 500.