SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2009-19

On a numerical and graphical technique for evaluating some models involving rational expectations

Søren Johansen and Anders Rygh Swensen

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

On a numerical and graphical technique for evaluating some models involving rational expectations

Søren Johansen, *
University of Copenhagen and CREATES, University of Aarhus †
and
Anders Rygh Swensen
University of Oslo and Statistics Norway

Abstract

Campbell and Shiller (1987) proposed a graphical technique for the present value model which consists of plotting the spread and theoretical spread as calculated from the cointegrated vector autoregressive model. We extend these techniques to a number of rational expectation models and give a general definition of spread and theoretical spread. The main results are the asymptotic distributions of the variance ratio, noise ratio, and correlation between the estimated spread and theoretical spreads. We derive sup tests for the recursively calculated quantities. Finally we apply the methods to two previous studies by Campbell and Shiller (1987) and Engsted (2002).

Classification: C32.
Keywords: VAR models, cointegration, rational expectations.

*Corresponding author: Søren Johansen, Department of Economics, University of Copenhagen, Studiestræde 6, DK-1455 Copenhagen K, Denmark

†Support from Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation is gratefully acknowledged by the first author.

1 Introduction

Vector autoregressive (VAR) models have been applied in several ways to analyze various forms of present value models. The most immediate one is to incorporate the restrictions entailed by the present value models to estimate the parameters of the VAR model to form a basis for statistical testing of the model.

Use of all purpose tests like likelihood ratio and Wald tests often leads to rejection of the restrictions. A natural question is then whether the rejection is due to essential parts of the present value models or is due to features of less importance. Campbell and Shiller (1987) proposed a graphical technique which addresses this issue in a present value model. As a special case of this let $\{X_t\} = (Y_t, y_t)'$ be a 2-dimensional time series consisting of the stock price at the end of period t and of the dividend paid during the period t. The present value model entails that the stock price can be expressed as a discounted sum of the expected future dividends given the information presently available i.e.

$$Y_t = \sum_{i=1}^{\infty} \delta^i E_t[y_{t+i}] + c, \tag{1}$$

which implies

$$Y_t = \delta E_t[Y_{t+1} + y_{t+1}] + c(1-\delta). \tag{2}$$

Note that if the information set contains more variables than the stock prices and dividends, the conditional expectations may involve these variables. Also, note that in Campbell and Shiller (1987) $c = 0$ for the present value model of stock prices. We find it useful for the discussion to allow for a general value of c.

Equation (2) shows that if both Y_t and y_t are $I(1)$ variables, $S_t^1 = Y_t - \frac{\delta}{1-\delta} y_t$ is stationary and we find the representation from (1)

$$Y_t - \frac{\delta}{1-\delta} y_t = \frac{1}{1-\delta} \sum_{i=1}^{\infty} \delta^i E_t[\Delta y_{t+i}] + c. \tag{3}$$

Campbell and Shiller (1987) denoted $S_t^1 = Y_t - \frac{\delta}{1-\delta} y_t$ the *(actual) spread*, and defined the *theoretical spread* as $S_t^2 = \frac{1}{1-\delta} \sum_{i=1}^{\infty} \delta^i E_t[\Delta y_{t+i}]$, so that the present value model can be expressed as,

$$S_t^1 = S_t^2 + c. \tag{4}$$

Alternatively, using (1) one may express the spread as

$$Y_t - \frac{\delta}{1-\delta} y_t = \frac{\delta}{1-\delta} E_t[\Delta Y_{t+1} + \Delta y_{t+1}] + c, \tag{5}$$

which says that the spread is linear in the optimal forecast of the change in the sum of the stock price and the dividend. Also (5) may be written in the form (4) by defining the theoretical spread as $S_t^3 = \frac{\delta}{1-\delta} E_t[\Delta Y_{t+1} + \Delta y_{t+1}]$.

The basic idea of the procedure suggested by Campbell and Shiller (1987), is to compare an estimator of the actual spread

$$S_t^1 = Y_t - \frac{\delta}{1-\delta} y_t,$$

and an estimator of the theoretical spread, S_t^2, that is, a forecast of a weighted average of the future change in dividends

$$S_t^2 = \frac{1}{1-\delta} \sum_{i=1}^{\infty} \delta^i E_t[\Delta y_{t+i}]. \tag{6}$$

Also one could consider forecasts of

$$S_t^3 = \frac{\delta}{1-\delta} E_t[\Delta Y_{t+1} + \Delta y_{t+1}]. \tag{7}$$

This seems as a sensible thing to do. Both the spread and the theoretical spread have intrinsic meaning. If the restrictions implied by the present value model are valid, both versions estimate the same thing. This idea is implemented by Campbell and Shiller in a number of steps

- Estimate the cointegrating relation $Y_t - \frac{\delta}{1-\delta} y_t$ by regression.
- Transform the data to the stationary variables $\hat{S}_t^1 = Y_t - \frac{\hat{\delta}}{1-\hat{\delta}} y_t$ and Δy_t.
- Fit a VAR model to the demeaned stationary variables $(\hat{S}_t^1, \Delta y_t)$.
- Calculate the forecast of $E_t[\Delta y_{t+i}]$ from the fitted model and calculate $\hat{S}_t^2$ from (6).
- Compare $\hat{S}_t^1$ and $\hat{S}_t^2$, e.g. by plotting them in the same diagram and compute statistics as correlation or variance ratio.

There are several problems with this procedure

1) The three first steps are in fact a two-step procedure which ignores that there is a relation between the original observations and the transformed series. In fact, the parameters of the stationary VAR model for the transformed series are functions of the parameters of the reduced rank VAR of the original observations.

2) There may be several possibilities for transforming to a stationary system, e.g. Campbell and Shiller (1987, p. 1067) mention that also $(\hat{S}_t^1, \Delta Y_t)$ can be used for fitting a VAR model. These alternatives may result in different fitted stationary models, and hence in different forecasts.

3) There are other definitions of a theoretical spread which can be used, as we have seen, i.e. $\hat{S}_t^3$.

4) For systems containing more than just two variables, there may be more stationary linear combinations of the variables than the one described by the spread, so the spread need not identify the cointegrating relations.

5) It is not clear what is the appropriate framework for evaluating the uncertainty of the forecasts, the original $I(1)$ system or the stationary VAR model for the transformed data.

We suggest in the present note that this implementation of the basic idea can be improved by conducting the analysis in a cointegrated VAR model of the originals series. There has recently been some work indicating that such results would be useful. Kurmann (2005) compared measures of theoretical and observed inflation in a New Keynesian pricing model following Campbell and Shiller's approach. Carriero et al (2006) compared spread between long- and short-term interes rates using similar ideas.

In the next section we define the model, show how the statistics of interest can be derived and consider some potential applications. In section 3 the asymptotic distribution of the actual spread and of the correlation and variance ratios are derived. In the last section the results using the procedures we propose are compared to those of two previous studies by Campbell and Shiller (1987) and Engsted (2002).

We use the following notation. If a $p \times r$ matrix α, where $r \leq p$, has full rank, $\alpha_\perp$ denotes a $p \times (p-r)$ matrix of full rank such that $\alpha'_\perp \alpha = 0$. The matrix $\alpha(\alpha'\alpha)^{-1}$ is defined as $\bar{\alpha}$, so that $\alpha'\bar{\alpha} = I_r$ and $\bar{\alpha}\alpha'$ is a projection matrix, $vec(\alpha)$ denotes the vector consisting of the stacked columns of α, and $\otimes$ is the Kronecker product defined as $C \otimes D = \{C_{ij}D\}$. Convergence in probability is denoted by $\xrightarrow{P}$ and $\Longrightarrow$ means convergence in distribution.

2 The statistical model

2.1 The statistical model

Assume that the $p-$dimensional data vector is generated by the cointegrated VAR

$$\Delta X_t = \alpha(\beta' X_{t-1} + \kappa_1 t) + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{k-1}\Delta X_{t-k+1} + \mu_0 + \varepsilon_t \tag{8}$$

where α and β are $p \times r$ matrices and that the errors ε_t are independent multivariate Gaussian variables with mean zero and covariance matrix Ω. Let $\beta^* = (\beta', \kappa_1)'$ and $X^*_{t-1} = (X'_{t-1}, t)'$.

For the asymptotic analysis we also assume that the process is $I(1)$, which means that the characteristic polynomial corresponding to model (8) has all zeros outside the unit

circle or exactly at 1. Furthermore, the matrix $\alpha'_\perp \Gamma \beta_\perp$, where $\Gamma = I_p - \Gamma_1 - \cdots - \Gamma_{k-1}$, has full rank $p - r$. Under these assumptions, we have the representation

$$\begin{aligned} X_t &= C \sum_{i=1}^{t} (\varepsilon_i + \mu_0) + \sum_{i=0}^{\infty} C_i (\varepsilon_{t-i} + \alpha \kappa_1' (t-i) + \mu_0) + A_0 \\ &= C \sum_{i=1}^{t} \varepsilon_i + \xi t + \xi_0 + Y_t + A_0 \end{aligned} \tag{9}$$

where Y_t is stationary, A_0 depends on initial values so that $\beta' A_0 = 0$, and the matrix C is given by

$$C = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp,$$

see e.g. Johansen (1995a). In order to forecast the process, it is convenient to consider model (8) formulated in the stationary companion form. From (8) it follows that

$$\begin{aligned} \beta^{*\prime} X_t^* &= \beta' X_t + \kappa_1 (t+1) = \beta' X_{t-1} + \kappa_1 (t+1) + \beta' \alpha (\beta' X_{t-1} + \kappa_1 t) \\ &\quad + \beta' \Gamma_1 \Delta X_{t-1} + \cdots + \beta' \Gamma_{k-1} \Delta X_{t-k+1} + \beta' \mu_0 + \beta' \varepsilon_t \\ &= (\beta' \alpha + I_r) \beta^{*\prime} X_{t-1}^* + \kappa_1 + \beta' \Gamma_1 \Delta X_{t-1} + \cdots + \beta' \Gamma_{k-1} \Delta X_{t-k+1} + \beta' \mu_0 + \beta' \varepsilon_t. \end{aligned} \tag{10}$$

We define the stacked stationary process

$$Z_{t-1} = Z_{t-1}(\beta^*) = (X_{t-1}^{*\prime} \beta^*, \Delta X_{t-1}', \cdots, \Delta X_{t-k+1}')'.$$

of dimension $l = r + p(k-1)$.

For $k > 1$, Z_t satisfies the $AR(1)$ equation

$$Z_t = A Z_{t-1} + \underline{\mu} + Q \varepsilon_t, \tag{11}$$

where the $l \times l$ matrix A and the $l \times p$ matrix Q are given by

$$A = \begin{pmatrix} \beta'\alpha + I_r & \beta'\Gamma_1 & \cdots & & \beta'\Gamma_{k-1} \\ \alpha & \Gamma_1 & \cdots & & \Gamma_{k-1} \\ 0 & I_p & & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & I_p & 0 \end{pmatrix}, Q = \begin{pmatrix} \beta' \\ I_p \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{12}$$

and $\underline{\mu} = Q\mu_0 + (\kappa_1', 0 \ldots, 0)'$. Because (11), regarded as a difference equation, has a stationary solution, all eigenvalues of A must have modulus less than one. For $k = 1$, the autoregressive equation can be expressed in $Z_t = \beta^{*\prime} X_t^*$, and $\beta'\alpha + I_r$. The autoregressive formulation implies that

$$\begin{aligned} E_t(Z_{t+i}) &= A^i Z_t + \sum_{j=0}^{i-1} A^j \underline{\mu} = A^i Z_t + B_i \underline{\mu}, \\ B_i &= (I_l - A^i)(I_l - A)^{-1}. \end{aligned}$$

From the conditional expectations $E_t(Z_{t+i})$ we can now pick out the conditional expectations of the differences, on which forecasts can be based, because

$$E_t(\Delta X_{t+i}) = (0_{p\times r}, I_p, 0_{p\times(l-r-p)})E_t(Z_{t+i}) = d'[A^i Z_t + B_i\underline{\mu}]. \tag{13}$$

For $k = 1$, the formula is easier as $E_t(\Delta X_{t+i}) = E_t(\alpha\beta^{*\prime}X^*_{t+i-1}) + \mu_0$ and $E_t(\beta^{*\prime}X^*_{t+i}) = (I_r + \beta'\alpha)E_t(\beta^{*\prime}X^*_{t+i-1}) + \kappa_1 + \beta'\mu_0$. Thus

$$E_t(\Delta X_{t+i}) = \alpha(I_r + \beta'\alpha)^{i-1}\beta^{*\prime}X^*_t - \alpha(\beta'\alpha)^{-1}[I_r - (I_r + \beta'\alpha)^{i-1}](\beta'\mu_0 + \kappa_1) + \mu_0.$$

For technical reasons it is convenient to use, in this case also, an autoregressive representation where the vector Z_{t-1} contains the differences ΔX_{t-1}. For $k = 1$ the matrix A will then be singular and equal to

$$A = \begin{pmatrix} \beta'\alpha + I_r & 0 \\ \alpha & 0 \end{pmatrix}. \tag{14}$$

Unless otherwise stated, when $k = 1$, A will be given by (14), $l = r + p$ and $Z_{t-1} = (X^{*\prime}_{t-1}\beta^*, \Delta X'_{t-1})'$.

2.2 A general definition of actual and theoretical spreads

In the statistical model it is convenient to define the spreads in terms of the stationary cointegrating relations. Thus we define in general the actual spread as

$$\beta^*_1(\upsilon)'X^*_t = \beta_1(\upsilon)'X_t + \kappa_1 t = a'Z_t(\beta^*). \tag{15}$$

When the spread is normed on the first element, $a = (1, 0, \ldots, 0) \in R^l$, otherwise $a = (1/\beta_{11}, 0, \ldots, 0) \in R^l$. If $\kappa_1 = 0$ we use $\beta_1(\upsilon)'X_t = a'Z_t(\beta)$, and if further $\mu = \alpha\kappa_0$ we use $\beta^*_1(\upsilon)'X^*_t = \beta_1(\upsilon)'X_t + \kappa_0 = a'Z_t(\beta^*)$.

The spread $\beta^*_1(\upsilon)$ is linear in υ, which is a vector of $m_1 \leq p + 1 - r$ parameters describing the spread. Then β^*_1 may be written

$$\beta^*_1 = h^*_1 + H^*_1\upsilon_1. \tag{16}$$

where h^*_1 is a $(p + 1) \times 1$ vector and H^*_1 is $(p + 1) \times m_1$ matrix.

We assume that the other $r - 1$ columns of β^* are *just* identified by linear restrictions and a normalization, which we express as

$$\beta^*_i = h^*_i + H^*_i\upsilon_i. \tag{17}$$

Here h^*_i is a $(p + 1) \times 1$ vector and H^*_i is $(p + 1) \times m_i = (p + 1) \times (p + 1 - r)$, matrix so that β^*_i has $m_i = p + 1 - r$ free parameters $i = 2, \ldots, r$. Also let $H^*_i = (H'_{1i}, H'_{2i})'$ and assume finally that all columns of β are identified.

A theoretical spread will be used in a wider sense than in Campbell and Shiller (1987). Thus both the formulations (1) and (5) may be taking as starting point, and it turns out in the examples below that the theoretical spread can written in the form

$$b'[f(A)Z_t(\beta^*) + g(A)\underline{\mu}] \tag{18}$$

where b is a suitably defined vector, and $f(A)$ and $g(A)$ are suitably defined matrices. Strictly speaking f and g also depend on the parameters in the cointegrating relations, but the estimators of these are super-consistent, so that we have suppressed them in the notation. All the estimators necessary to compute $\hat{A}$ and $\hat{\underline{\mu}}$ i.e. $\hat{\upsilon}, \hat{\alpha}, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}, \hat{\mu}_0, \hat{\kappa}_1$ are the maximum likelihood estimator in the model (8), conditioned on the initial observations $X_{-k+1}, \cdots, X_0$.

Example 1: the present value model. Consider again the present value model with $X_t = (Y_t, y_t, z_t')'$, where Y_t and y_t are defined as before and z_t are additional variables of interest to be included in the information set. Let a be the l-dimensional vector where only the first element is non-zero and equal to one. Then the actual spread, i.e. S_t^1 from the introduction, can be expressed as

$$S_t^1 = Y_t - \frac{\delta}{1-\delta} y_t = a' Z_t,$$

and $\beta_1^* = (1, -\frac{\delta}{1-\delta}, 0, \ldots, 0)'$, is assumed identified by the zero restrictions. In this case there is only one parameter, $\upsilon = \delta/(1-\delta)$, to be estimated, and $m_1 = 1$.

As for the theoretical spread, i.e. S_t^2, as in the introduction,

$$\begin{aligned}\sum_{i=1}^{\infty} \delta^i E_t[Z_{t+i}] &= \delta A(I_l - \delta A)^{-1} Z_t + \sum_{i=1}^{\infty} \delta^i \sum_{j=1}^{i} A^{j-1} \underline{\mu} \\ &= \delta A(I_l - \delta A)^{-1} Z_t + \frac{\delta}{1-\delta}(I_l - \delta A)^{-1} \underline{\mu}.\end{aligned}$$

Pre-multiplying Z_t by $b' = e_2' d'$, where d is defined in (13) and e_i is the p-dimensional vector where only the i'th element is non-zero and equal to one, picks out the first difference series for the dividends. Taking

$$f(A) = \frac{\delta}{1-\delta} A(I_l - \delta A)^{-1}, \; g(A) = \frac{\delta}{(1-\delta)^2}(I_l - \delta A)^{-1},$$

the theoretical spread may be written on the form (18).

An alternative formulation is to express the spread as the optimal forecast of the change in the sum of the stock price and dividend, i.e. as S_t^3 in (7). Then the theoretical spread, which is also of the form (18), will be

$$S_t^3 = \frac{\delta}{1-\delta}(1, 1, 0, \ldots, 0) d' (\hat{A} Z_t(\beta^*) + \underline{\mu})$$

where

$$f(A) = \frac{\delta}{1-\delta}A,\ g(A) = \frac{\delta}{1-\delta}I_l,\ b' = (1, 1, 0, \ldots, 0)d'. \qquad \blacksquare$$

Example 2: uncovered interest parity. If $X_t = (i_{1,t}, i_{2,t}, d_t)'$, where $i_{1,t}$ and $i_{2,t}$ are domestic and foreign interest rates respectively, and d_t is the depreciation of own currency. The uncovered interest parity is defined as

$$i_{1,t} - i_{2,t} = E_t[d_{t+1}],$$

which can also be expressed as

$$i_{1,t} - i_{2,t} - d_t = E_t[\Delta d_{t+1}].$$

In this case it is natural to consider the forecast of the right hand side given by

$$\hat{S}_t^2 = e_3' d' (\hat{A} Z_t(\hat{\beta}^*) + \hat{\underline{\mu}}),$$

so that

$$f(A) = A,\ g(A) = 1,\ b' = e_3' d'. \blacksquare$$

Example 3: term structure of interest rate. Consider next the term structure of interest rate where $R_t^{(n)}$ is an n-period discount yield or interest rate. We focus on periods of three and one months. The linearized form of the expectation hypotheses of the term structure (EHTS) can be expressed as

$$R_t^{(3)} = \frac{1}{3}(R_t^{(1)} + E_t[R_{t+1}^{(1)}] + E_t[R_{t+2}^{(1)}]) + c \tag{19}$$

where c is a risk premium. Thus, the EHTS implies that the actual interest spread

$$S_t^1 = R_t^{(3)} - R_t^{(1)} = \frac{1}{3}E_t[\Delta R_{t+2}^{(1)}] + \frac{2}{3}E_t[\Delta R_{t+1}^{(1)}] + c$$

is a cointegrating relations if $(R_t^{(3)}, R_t^{(1)})$ is $I(1)$.

Therefore, if we let $X_t = (R_t^{(3)}, R_t^{(1)}, z_t')'$, where the elements of z_t are additional variables of interest, the EHTS model fit into the framework described above. Obvious candidates as elements of z_t are other interest rates of different periods. Then more elaborate EHTS than (19) may also be investigated.

The actual spread is directly observed, so $\hat{S}_t^1 = S_t^1$, $\beta_1 = (1, -1, 0, \ldots, 0)'$ and there are no parameters to be estimated in the actual spread, i.e. $m_1 = 0$. A reduced rank VAR model (8) with a restricted constant term, $\mu_0 = \alpha\kappa_0'$, appears to be most natural, as linear trends are unlikely to be present in interest data. The stacked process satisfies the equation

$$Z_t = AZ_{t-1} + Q\alpha\kappa_0' + Q\varepsilon_t,$$

and gives the conditional expectations

$$E_t[\Delta R^{(1)}_{t+1}] = e_2' E_t[d' Z_{t+1}] = e_2' d'(AZ_t + Q\alpha\kappa_0'),$$
$$E_t[\Delta R^{(1)}_{t+2}] = e_2' \hat{E}_t[d' Z_{t+2}] = e_2' d'(A^2 Z_t + AQ\alpha\kappa_0' + Q\alpha\kappa_0').$$

Thus in this context the theoretical spread is

$$S_t^2 = \frac{1}{3} e_2' d'[A^2 Z_t + 2AZ_t + AQ\alpha\kappa_0' + 3Q\alpha\kappa_0'].$$

In the case

$$f(A) = \frac{1}{3}(A^2 + 2A),\ g(A) = \frac{1}{3}(A + 3I_l),\ b' = e_2' d'. \qquad \blacksquare$$

3 Asymptotic distributions of spreads, correlation, variance ratio, and noise ratio

We consider the asymptotic distribution of the spreads and of numerical summaries used to describe them such as correlation, variance ratio and noise ratio. A process of considerable interest is the difference of the estimators of the spreads, $T^{1/2}(\hat{S}_t^1 - \hat{S}_t^2)$. It turns out that the limit does not exist. Therefore, we have to analyze it by considering some functionals like those mentioned, correlation, variance ratio and noise ratio.

3.1 The spreads

Both the actual spread, $S_t^1 = \beta_1^*(\upsilon)' X_t^* = a' Z_t$, where a is the first unit vector, and the theoretical spread $S_t^2 = b'[f(A)Z_t + g(A)\underline{\mu})]$, where b is a suitable selection vector, can be expressed in terms of Z_t. For $S_t = (S_t^1, S_t^2)'$ we define ψ and η by

$$\bar{S}_T = \begin{pmatrix} a' \\ b' f(A) \end{pmatrix} \bar{Z}_T + \begin{pmatrix} 0 \\ b' g(A)\underline{\mu} \end{pmatrix} = \psi \bar{Z}_T + \eta, \tag{20}$$

and let $\mu_Z = E[Z_t]$, $\mu_S = \psi\mu_Z + \eta$ and $\Sigma_S = \psi\Sigma_Z\psi'$ where Σ_Z is the covariance matrix of Z_t. Then, let

$$U_T = \frac{1}{T}\sum_{t=1}^{T}(S_t - \mu_S)(S_t - \mu_S)' = \psi \frac{1}{T}\sum_{t=1}^{T}(Z_t - \mu_Z)(Z_t - \mu_Z)'\psi'. \tag{21}$$

Denote by $\widehat{\bar{S}}_T, \hat{U}_T$ the estimates we get by replacing all parameters by their maximum likelihood estimates. In the appendix the following proposition is proved.

Proposition 1 *Let f and g be continuously differentiable and define the derivative $F = \partial vec(f(A))/\partial(vec(A))'$. Then*

i)

$$\hat{\bar{S}}_T \xrightarrow{P} \mu_S \tag{22}$$

$$\hat{U}_T \xrightarrow{P} \Sigma_S. \tag{23}$$

ii) On $D^4[0,1]$*, we find*

$$u\sqrt{T}vec(\hat{U}_{[Tu]} - \hat{U}_T) \Longrightarrow W_U(u) - uW_U(1)$$

where W_U *is Brownian motion with variance* Σ_U*, which is found as the variance of*

$$[\{(\zeta \otimes \psi\Sigma_Z)K_{ll} + (\psi\Sigma_Z \otimes \zeta)\}F(\Sigma_Z^{-1} \otimes I), (\psi \otimes \psi)]V. \tag{24}$$

where ψ *is defined in (20) and* $\zeta = (0_{l\times 1}, b)'$*. The variance of* $V = (V_1', V_2')'$ *is given by (27) in Lemma 2 in the appendix.*

When $k = 1$*,* Σ_Z^{-1} *in (24) shall be substituted with* $(I_r\ 0)'((I_r\ 0)\Sigma_Z(I_r\ 0)')^{-1}(I_r\ 0)$*.*

Remark 1. It is worth emphasizing that the asymptotic variance of $vec(\hat{U}_T)$ consists of two parts. On the one hand, the term $(\psi \otimes \psi)V_2$, which depends on the coefficients of the VAR only through the value of $f(A)$. On the other hand, a component which depends on the linearization of $f(A)$ with respect to A, and which will take into account the variation in the estimated coefficients in A.

There are also other random variables whose distributions are of interest. As shown in Lemma 3 in the appendix the asymptotic distribution of $T^{1/2}(\hat{S}^1_{[Tu]} - S^1_{[Tu]}) = T^{1/2}a'(\hat{Z}_{[Tu]} - Z_{[Tu]}) = T^{1/2}(\beta_1^*(\hat{\upsilon}) - \beta_1^*(\upsilon))'X^*_{[Tu]}$ is mixed Gaussian. But the asymptotic distribution of $T^{1/2}(\hat{S}^2_{[Tu]} - S^2_{[Tu]})$ is rather involved, as we shall now see. We find that

$$\begin{aligned} T^{1/2}(\hat{S}^2_{[Tu]} - S^2_{[Tu]}) &= T^{1/2}[b'f(\hat{A})Z_{[Tu]}(\hat{\beta^*}) - b'f(A)Z_{[Tu]}(\beta^*) + g(\hat{A})\hat{\mu} - g(A)\mu] \\ &= b'f(\hat{A})T^{1/2}[Z_{[Tu]}(\hat{\beta^*}) - Z_{[Tu]}(\beta^*)] + b'T^{1/2}[f(\hat{A}) - f(A)]Z_{[Tu]}(\beta^*) \\ &\quad - b'T^{1/2}[g(\hat{A})\underline{\hat{\mu}} - g(A)\underline{\mu}] + O_P(T^{-1}). \end{aligned}$$

The first term converges in distribution, as we have pointed out, and the last term converges by an application of the delta method. The second term is more complicated because although the factor $T^{1/2}[f(\hat{A}) - f(A)]$ converges in distribution by an application of the delta method, the process $Z_{[Tu]}(\beta^*)$ does not converge to anything as a process in $D^l[0,1]$.

The same remark holds for the difference of the estimated actual spread and the estimated theoretical spread: $\hat{S}^1_{[Tu]} - \hat{S}^2_{[Tu]}$, which of course is the most interesting process to analyze.

As a consequence some functionals of the processes $\hat{S}^1_{[Tu]}$ and $\hat{S}^2_{[Tu]}$ for which we can find the asymptotic properties, will be considered in the next section.

3.2 Correlation, variance ratio and noise ratio

The dependency of the movements of the estimates of the actual and theoretical spreads is often described by the correlation, the variance ratio and the noise ratio of the two series, defined as

$$\begin{aligned} \rho &= Cov(S_t^1, S_t^2)/\sqrt{Var(S_t^1)Var(S_t^2)}, \\ vr &= Var(S_t^1)/Var(S_t^2), \\ nr &= Var(S_t^1 - S_t^2)/Var(S_t^1). \end{aligned}$$

We give below the asymptotic distributions of these quantities estimated recursively on the interval $1 \leq t \leq [Tu]$, that is as processes on $D[0,1]$. We find that for ρ, say, we get

$$\Sigma_\rho^{-1/2} u T^{1/2}(\hat{\rho}_{[Tu]} - \hat{\rho}_T) \Longrightarrow B(u) - uB(1)$$

so that sup tests can be applied. Here $\Sigma_\rho^{-1/2}$ is the asymptotic variance of $\hat{\rho}_T$, as given in Proposition 2 below, and $B(u)$ is a standard Brownian motion. Similar results hold for the other quantities.

It follows from Proposition 1 that the estimators are consistent, so that

$$\hat{\rho}_T \xrightarrow{P} \rho = \frac{a'\Sigma_Z f(A)'b}{\sqrt{a'\Sigma_Z a b' f(A)\Sigma_Z f(A)'b}} \tag{25}$$

$$\widehat{vr}_T \xrightarrow{P} vr = \frac{a'\Sigma_Z a}{b' f(A)\Sigma_Z f(A)'b}$$

$$\widehat{nr}_T \xrightarrow{P} nr = \frac{a'\Sigma_Z a + b' f(A)\Sigma_Z f(A)'b - 2a'\Sigma_Z f(A)'b}{a'\Sigma_Z a} \tag{26}$$

where Σ_Z is the covariance matrix of Z_t.

In the special case of $k = 1$ and $r = 1$, there is only one cointegration vector, and both estimated spreads can be expressed as linear functions of $\hat{\beta}' X_t^*$, which is scalar. This can be seen from the discussion at the end of section 2.1, because the conditional expectations of the differences are functions of $\beta' X_t^*$ only. One consequence is that both the empirical, $\hat{\rho}$, and population correlation, ρ equal 1. Another is that the three dimensional asymptotic distribution considered in Proposition 2 must be degenerate.

By a straightforward application of the delta method we get the following result from Proposition 1.

Proposition 2 *Let matrix Σ_u be the asymptotic 3×3 variance of $T^{1/2}(\hat{U}_{T11}, \hat{U}_{T12}, \hat{U}_{T22})$, which can be found as a sub-matrix of Σ_U in Proposition 1. Then*

i)

$$uT^{1/2}(\hat{\rho}_{[Tu]} - \hat{\rho}_T) \Longrightarrow W_\rho(u) - uW_\rho(1),$$

where the variance of W_ρ *is* $\Sigma_\rho = \frac{1}{4}\rho^2 c_1' \Sigma_u c_1$ *for*

$$c_1 = [\frac{1}{a'\Sigma_Z a}, \frac{-2}{a'\Sigma_Z f(A)'b}, \frac{1}{b'f(A)\Sigma_Z f(A)'b}]'.$$

ii)

$$uT^{1/2}(\widehat{vr}_{[Tu]} - \widehat{vr}_T) \Longrightarrow W_{vr}(u) - uW_{vr}(1),$$

where W_{vr} *has variance* $\Sigma_{vr} = c_2' \Sigma_u c_2$ *for*

$$c_2 = (b'f(A)\Sigma_Z f(A)'b)^{-2}[b'f(A)\Sigma_Z f(A)'b, 0, -a'\Sigma_Z a]'.$$

iii)

$$uT^{1/2}(\widehat{nr}_{[Tu]} - \widehat{nr}_T) \Longrightarrow W_{nr}(u) - uW_{nr}(1),$$

where W_{nr} *has variance* $\Sigma_{nr} = c_3' \Sigma_u c_3$ *for*

$$c_3 = (a'\Sigma_Z a)^{-2}[-b'f(A)\Sigma_Z f(A)'b + 2a'\Sigma_Z f(A)'b, -2a'\Sigma_Z a, a'\Sigma_Z a]'.$$

Example 1, continued. Here $f(A) = \frac{\delta}{(1-\delta)} A(I_l - \delta A)^{-1} = \frac{1}{(1-\delta)}[-I_l + (I_l - \delta A)^{-1}]$ which shows that

$$\begin{aligned} f(\hat{A}) - f(A) &= \frac{1}{(1-\delta)}[(I_l - \delta\hat{A})^{-1} - (I_l - \delta A)^{-1}] \\ &= \frac{\delta}{(1-\delta)}(I_l - \delta A)^{-1}(\hat{A} - A)(I - \delta\hat{A})^{-1}. \end{aligned}$$

The linearization needed to compute the variance in (24) is therefore in this case,

$$\begin{aligned} vec\{[f(\hat{A}) - f(A)]\} &= \frac{\delta}{(1-\delta)}[(I_l - \delta\hat{A}')^{-1} \otimes (I_l - \delta A)^{-1}]vec((\hat{A} - A)) \\ &= \frac{\delta}{(1-\delta)}[(I_l - \delta A')^{-1} \otimes (I_l - \delta A)^{-1}]vec((\hat{A} - A)) + o_P(T^{-1/2}), \end{aligned}$$

and the estimated standard error of the correlation can be worked out as described above.

In addition to the *level* variance ratio based on $\hat{S}_t$ which works for general models of the form we consider, Campbell and Shiller (1987) also suggested an *innovation* variance ratio which is tailored to present value models. The variable

$$\xi_t^1 = S_t^1 - (\frac{1}{\delta})S_{t-1}^1 + \frac{\delta}{1-\delta}\Delta y_t$$

is a measure of the excess return on stock.

Now, define a similar quantity in terms of the theoretical spread

$$\xi_t^2 = S_t^2 - (\frac{1}{\delta})S_{t-1}^2 + \frac{\delta}{1-\delta}\Delta y_t.$$

After some algebra this is seen to equal

$$\xi_t^2 = \frac{1}{1-\delta}\sum_{i=0}^{\infty}\delta^i\{E_t[\Delta y_{t+i}] - E_{t-1}[\Delta y_{t+i}]\}$$

which can be interpreted as an innovation from time $t-1$ to time t of the expected present value of Δy_t. Let $\hat{\xi}_t = (\hat{\xi}_t^1, \hat{\xi}_t^2)'$ where $\hat{\xi}_t^1$ and $\hat{\xi}_t^2$ are the estimators of ξ_t^1 and ξ_t^2 respectively.

Then we may proceed as described for the level variance ratio to derive the asymptotic distribution of the innovation variance ratio because both innovations may be expressed as linear combinations of elements of Z_t and Z_{t-1}. ■

Example 2, continued. Then $f(A) = A$, so the asymptotic distribution follows from (24) with $F = I_{l^2}$. ■

Example 3, continued. In this case

$$f(\hat{A}) - f(A) = \frac{1}{3}\hat{A}^2 + \frac{2}{3}\hat{A} - \frac{1}{3}A^2 - \frac{2}{3}A$$
$$= (\frac{1}{3}(\hat{A}+A) + \frac{2}{3}I_l)(\hat{A}-A) = \frac{2}{3}(A+I_l)(\hat{A}-A) + o_P(T^{-1/2}),$$

so the asymptotic distribution of the correlation and the variance ratio can be worked out using (24) with $F = I_l \otimes \frac{2}{3}(A+I_l)$. ■

Remark 2. In order to take into account that the correlation is bounded by 1 in absolute value it is useful to consider the function $\frac{1}{2}\log(\frac{1+\hat{\rho}}{1-\hat{\rho}})$, with inverse tanh, which is asymptotically Gaussian with mean $\frac{1}{2}\log(\frac{1+\rho}{1-\rho})$ and variance $\frac{1}{4T}(\frac{\rho}{1-\rho^2})^2 c_1'\Sigma_u c_1$. An approximate 95% confidence interval has therefore bounds $\tanh[\frac{1}{2}\log(\frac{1+\hat{\rho}}{1-\hat{\rho}}) \pm \frac{1.96\hat{\rho}}{1-\hat{\rho}^2}\sqrt{\frac{1}{4T}c_1'\hat{\Sigma}_u c_1}]$.

Remark 3. Campbell and Shiller (1989) suggested computing the asymptotic distribution of the correlation and variance ratio using numerical differentiation with respect to the elements of the matrix A.

Remark 4. The arguments used to derive the asymptotic distributions relied on the fact that the parameters in the cointegration vectors are super-consistent, and hence can be treated as known. The situation where all these parameters actually are known, as in Examples 2 and 3, is therefore completely covered by the treatment above.

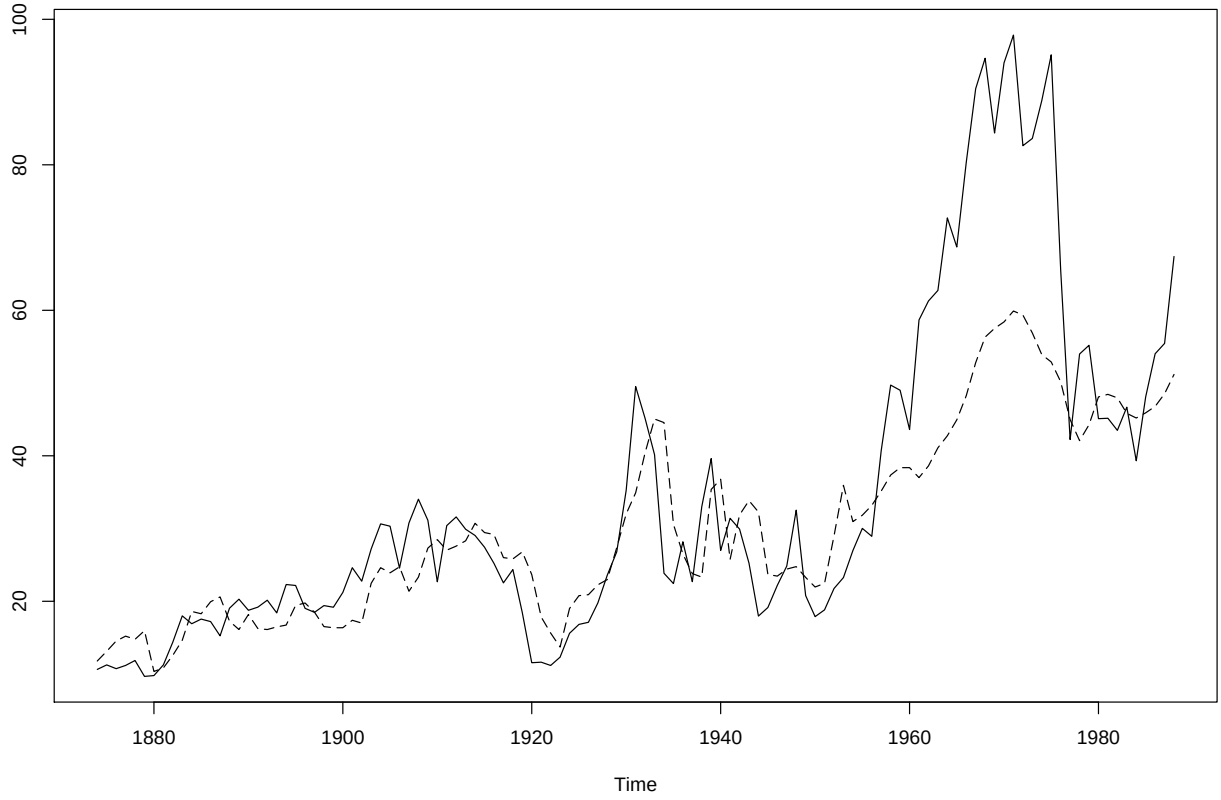

Figure 1: Real stock prices (solid line) and real dividends, multiplied by 20, (dashed line) of American stocks 1872-1986.

4 Some simple applications

We will compare the result of applying the methods proposed in this paper to the results reported in Campbell and Shiller (1987) and Engsted (2002). These studies are carried out in a bivariate setup so the situations are particularly simple, as pointed out in the introduction. The intention of the exercises is therefore to compare the traditional proposals and the modifications we suggest, and see how they differ. Both contain important lessons.

4.1 A present value model of Campbell and Shiller reconsidered

Campbell and Shiller used a VAR-model containing an unrestricted constant term to analyze series of U.S. stock prices and dividends for the years 1872-1986. The series are displayed in Figure 1. As explained in the introduction they fitted a VAR-model of lag

length two to two-dimensional series consisting of the estimated actual spread and the differenced dividend series after having removed the means. Let $X_t = (P_t, D_t)'$, where P_t is the stock price series, and D_t is the dividend series.

We fit a two-dimensional VAR of the form

$$\Delta X_t = \alpha\beta' X_{t-1} + \Gamma \Delta X_{t-1} + \mu + \varepsilon_t,$$

i.e. a model with an unrestricted constant term. A likelihood ratio test of the restrictions on the coefficients implied by the present value relations yields a p-value of 0.11, see Johansen and Swensen (1999).

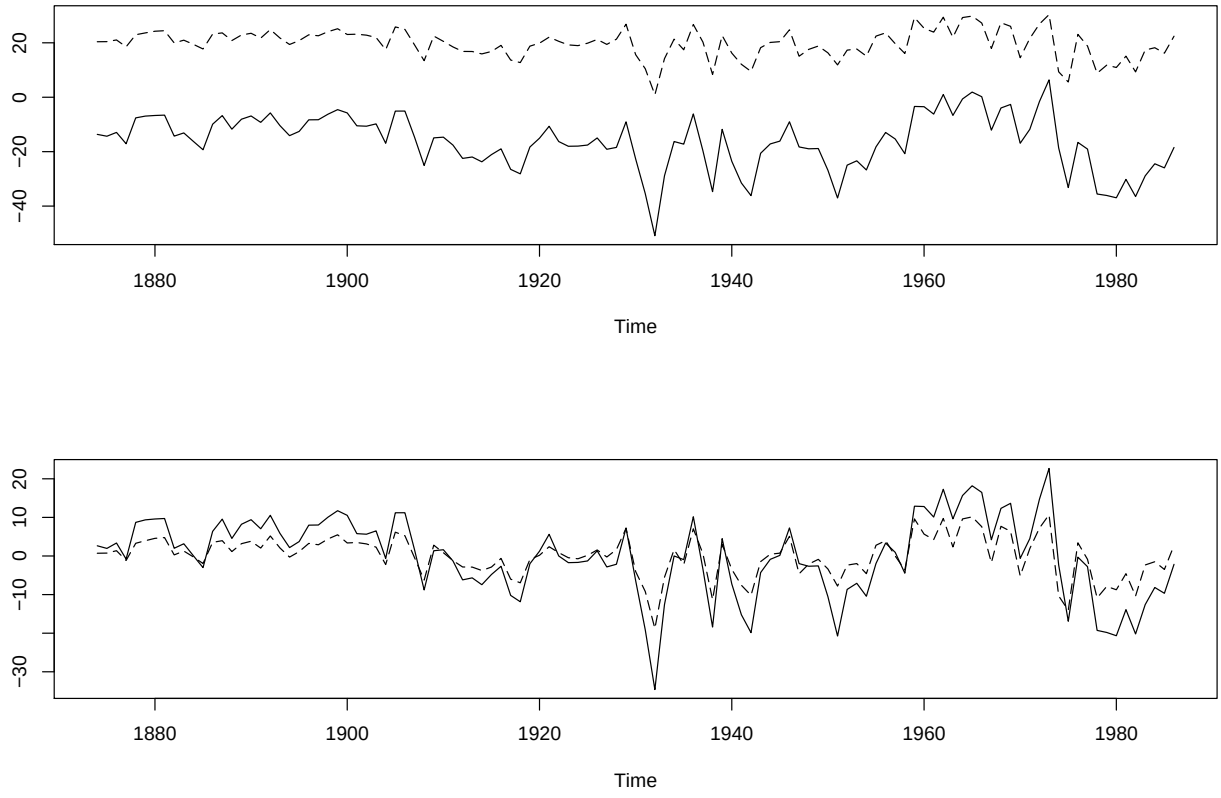

Figure 2: Plot of estimates of actual spread (solid line) and first theoretical spread, S_t^2, (dashed line) of U.S. stocks 1872-1986: upper panel as described in text, lower panel centered at the means.

The method for estimating the theoretical spread described in Section 2 can be used

Table 1: The matrix $\hat{A}$ for U.S. stocks (standard errors in parenthesis)

0.733	−0.170	−9.815
(0.076)	(0.115)	(4.055)
−0.118	0.160	−4.683
(0.071)	(0.107)	(3.780)
0.004	0.010	0.153
(0.001)	(0.002)	(0.076)

with $\kappa_1 = 0$ and $Z_{t-1}(\beta) = Z_{t-1} = (X'_{t-1}\beta, \Delta X'_{t-1})'$. Figure 2 shows plots of the actual spread and the estimated theoretical spread defined as estimated present value of the dividends, in the upper panel. This corresponds to estimates of both sides of (4) in the introduction, with $c = 0$. As one can see the levels of the two series are quite different. However, when series are centered at the same mean, corresponding to Campell and Shiller's (1987) demeaning, the result is similar to the corresponding one in Campbell and Shiller (1987). These results are not so surprising in light of the linear trend which is evident in the plot of the series. The transformation used by Campbell and Shiller (1987) removes this to a large extent, so the transformed series appear stationary. A further demeaning, before fitting a VAR-model without a constant, centers both spreads approximately at zero. However, the modified method we propose is based on a VAR-model with a constant. Due to the linear trend the differences of the dividends are mostly positive. Because the theoretical spread is a weighted sum of these differences, the estimates of the theoretical spread can also be expected to be positive, as is indeed the case, see the upper panel in Figure 2. Using the approach of Campbell and Shiller (1987) it is difficult to discover a non-zero c in formulas (1)-(5) when the series contain a trend, but using the modified approach such a feature appears immediately.

Table 2: Comparison of results by using the modified and traditional procedures.

	Modified procedure			Traditional procedure	
	estimate	est. std. err., two terms	est. std. err., one term	estimate	est. std. err.
$corr(S_T^1, S_T^2)$	0.94	0.11	0.10	0.911	0.207
$vr(S_T^1, S_T^2)$	3.60	2.37	2.31	4.786	5.380
$nr(S_T^1, S_T^2)$	0.29	0.25	0.24	-	-
$corr(S_T^1, S_T^3)$	-0.46	0.39	0.43	-	-
$vr(S_T^1, S_T^3)$	0.05	0.05	0.05	-	-
$nr(S_T^1, S_T^3)$	23.52	20.60	20.25	-	-
$100(1-\delta)/\delta$	2.98	-	-	3.2	-

In Figure 3 a similar exercise is done, but now defining the theoretical spread as linear in the forecast of the change in stock price and dividend, i.e as S_t^3 defined in (7) in the

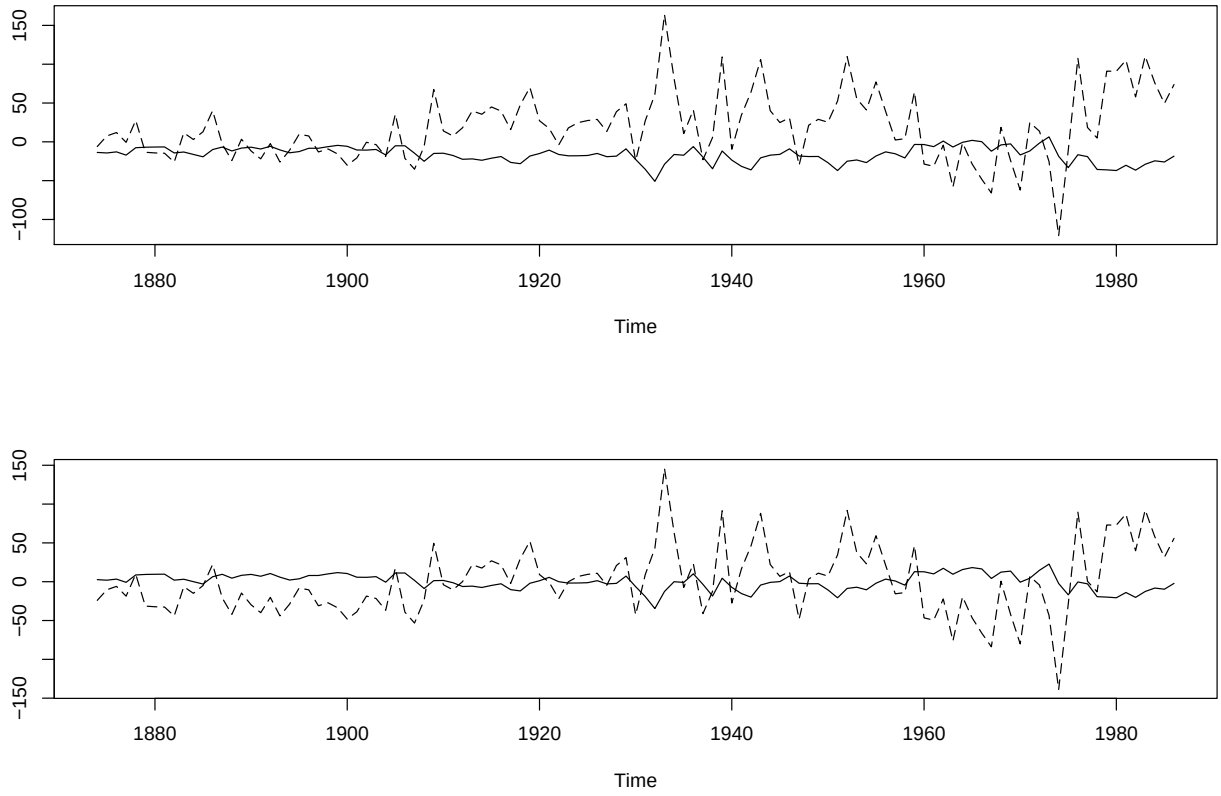

Figure 3: Plot of estimates of actual spread (solid line) and second theoretical spread, S_t^3, (dashed line) of U.S. stocks 1872-1986: upper panel as described in text, lower panel centered at the means.

introduction. In this case the two spreads differ much more. The discrepancy may not be so surprising if one looks at the estimated matrix $\hat{A}$, which can be found in Table 1. Remark the large, but not significant, negative value of the entry corresponding to the element $(1,2)$ in $\hat{\Gamma}$, -4.68. In the estimate of the first theoretical spread only forecasts of the change in dividends are used, which means that the last row of $\hat{A}(I-\hat{\delta}\hat{A})^{-1}$ is used for combining the elements of Z_t. With $\hat{\delta}=0.971$ this row equals (0.012,0.009,0.001). For the estimate of the second theoretical spread both forecasts of the stock prices and dividends are used. Hence, the two last rows in $\hat{A}$ are involved, which means that the coefficient -4.68 will have a direct influence on the estimated spread in this case.

It may be worth pointing out that Campbell and Shiller (1987) do not find the present value model for stock prices convincing either, see e.g. pp. 1083-1085. Hence, the discrepancy between the behavior of the two theoretical spreads supports their impression. Turning to the numerical summaries, the results from the modifications we propose are compared to those of Campbell and Shiller (1987) in the first and second row of Table 2. The two rightmost columns in Table 2 are taken from Campbell and Shiller (1987). The estimates of the standard errors from the asymptotic distribution are smaller than the estimates using numerical differentiation proposed by Campbell and Shiller (1989), regardless of whether one or two terms of (24) are included in the asymptotic variance. Actually, Campbell and Shiller (1989) treat the covariance matrix of Z_t as fixed. That corresponds to the situation where the asymptotic variance as given in (24) in Proposition 1, is calculated just from the first term in (24). When the covariance matrix is not considered as fixed an extra term, $(\psi\otimes\psi)V_2$, is necessary. As to the estimate of the variance ratio reported in the second row, remark the estimates of the standard errors are similar, regardless of whether one or two terms are included in the estimate of the variance. Also, note that the variance ratio is not significantly different from 1, and the noise ratio is not significantly different from zero. The results for the alternative theoretical spread defined by the forecast of the change in the sum of the one period stock price and dividend, i.e what is denoted S_t^3 in the introduction are displayed in rows four to six. They confirm the impression one gets from Figure 3.

In Figure 4 a recursive plot of the correlation between the estimate of the actual and the first theoretical spread is displayed. From Proposition 2 it follows that the asymptotic distribution is a Brownian bridge on $[0,1]$, whose maximum has a 95%-quantile equal to 1.36. There is therefore no indication that the correlation is nonconstant from a test based on the supremum of $\hat{\rho}_t, t=1874,\ldots,1986$.

4.2 A present value model for Danish stock prices

Consider next the example used by Engsted (2002) to illustrate various measures of fit for rational expectations models. The yearly bivariate time series $X_t=(P_t,D_t)'$ he considered consists of real stock prices, P_t, and dividends, D_t, for the period 1922-1996 and is displayed in Figure 4. He computed estimates of the actual and theoretical spread

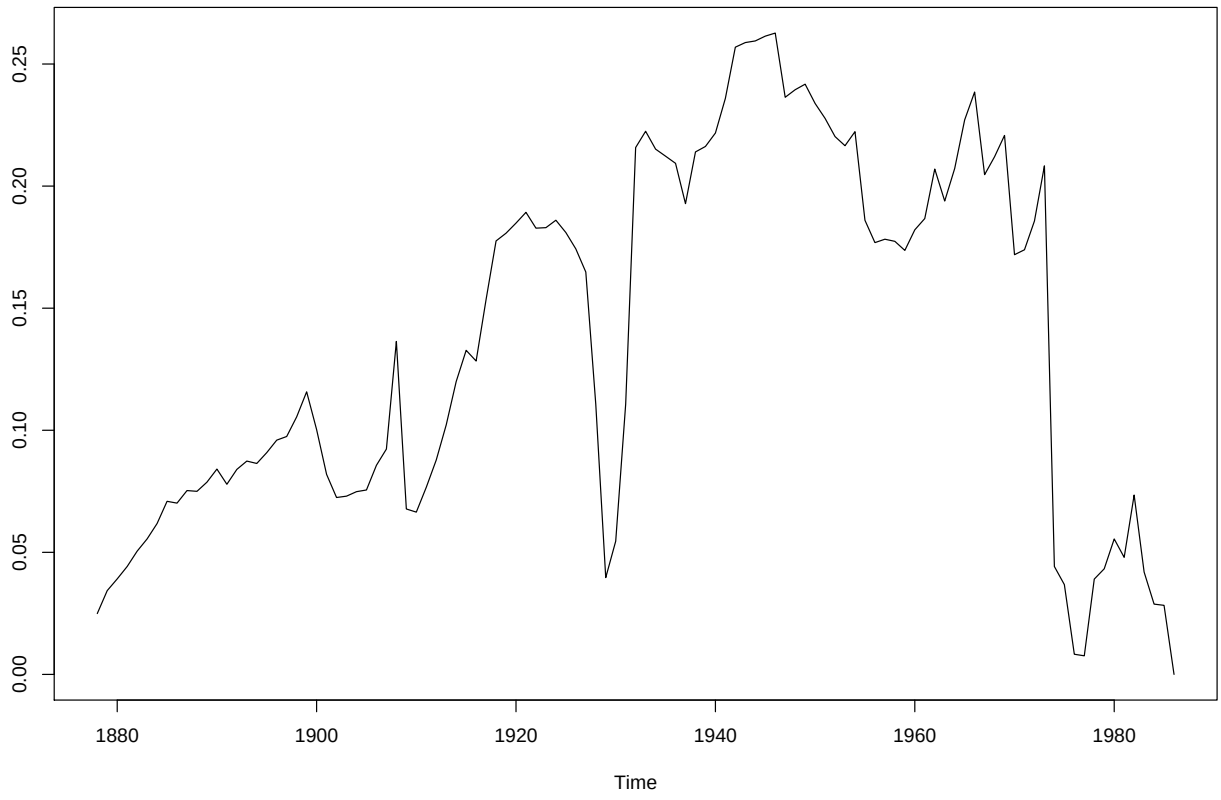

Figure 4: Recursive plot of $tT^{-1/2}\hat{\Sigma}_\rho^{-1/2}(\hat{\rho}_t - \hat{\rho}_T)$ where $\hat{\rho}_t = \widehat{\text{corr}}(S_t^1, S_t^2)$ of U.S. stocks 1872-1986.

by the procedure proposed by Campbell and Shiller (1987). The stock index P_t is a weighted portfolio measured at the end of year t, and D_t is the dividend paid during the same year. As in Engsted (2002) we used lag length 1 and fitted a reduced rank VAR-model of the form

$$\Delta X_t = \alpha\beta' X_{t-1} + \alpha\kappa_0 + \varepsilon_t.$$

The p-value of the likelihood ratio test of the restrictions on the coefficients implied by the present value relations is 0.21, see Johansen and Swensen (2004). The method for estimating the theoretical spread described in Section 2 can be applied with $\kappa_1 = 0, \mu_0 = \alpha\kappa_0$. The actual spread, S_t^1, is the normed cointegration vector $S_t^1 = P_t - \delta/(1-\delta)D_t + \kappa_0$ because $\beta_2/\beta_1 = -\delta/(1-\delta)$. Using the appropriate modifications described in Section 2,

Figure 5: Real stock prices (solid line) and real dividends (dashed line) of Danish stocks 1922-1996.

because the lag length k is equal t, we get

$$\Delta X_t = \alpha\beta^{*\prime}\begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix} + \varepsilon_t \text{ and } \beta^{*\prime}\begin{pmatrix} X_t \\ 1 \end{pmatrix} = (1+\beta'\alpha)\beta^{*\prime}\begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix} + \beta'\varepsilon_t$$

where $\beta^* = (\beta', \kappa_0)'$. Then

$$\sum_{i=1}^{\infty}\delta^i E_t[\Delta X_{t+i}] = \sum_{i=1}^{\infty}\delta^i\alpha[1+\beta'\alpha]^{i-1}\beta^{*\prime}\begin{pmatrix} X_t \\ 1 \end{pmatrix} = \delta\alpha[1-\delta(1+\beta'\alpha)]^{-1}\beta^{*\prime}\begin{pmatrix} X_t \\ 1 \end{pmatrix},$$

and the theoretical spread is

$$S_t^2 = \frac{1}{1-\delta}\sum_{i=1}^{\infty}\delta^i E_t[\Delta D_{t+i}] = \frac{\delta\alpha_2}{(1-\delta)(1-\delta(1+\beta'\alpha))}(P_t - \frac{\delta}{1-\delta}D_t + \kappa_0).$$

Table 3: Comparison of results for Danish stocks using the modified procedure with the results of Engsted

	Modified procedure		Engsted results
	estimate	sd	estimate
$corr(S_t^1, S_t^2)$	1.00	0.000	0.9999
$vr(S_t^1, S_t^2)$	3.35	1.882	3.38
$nr(S_t^1, S_t^2)$	0.205	0.139	0.210
$100(1-\delta)/\delta$	6.17	–	

Both estimated spreads are therefore linear functions of the estimated cointegration vector. Hence the correlation is 1, which may explain the correlation 0.999 obtained by Engsted (2003) after fitting a VAR model to $(\hat{S}_t^1, \Delta D_t)'$.

Also the alternative definition of the theoretical spread, i.e the one denoted by S_t^3 in the introduction, has this feature because

$$E_t \begin{pmatrix} \Delta P_{t+1} \\ \Delta D_{t+1} \end{pmatrix} = \alpha\beta^{*\prime} \begin{pmatrix} P_t \\ D_t \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (P_t - \frac{\delta}{1-\delta} D_t + \kappa_0),$$

so that $S_t^3 = E_t[\Delta P_{t+1} + \Delta D_{t+1}] = (\alpha_1 + \alpha_2)P_t - \delta/(1-\delta)D_t + \kappa_0$.

Figure 6, which displays the actual and first theoretical spread, is almost identical to the corresponding figure in Engsted (2002). The ratio between the theoretical and actual spread, which we denote as γ, is estimated as $\hat{\gamma} = (\hat{\delta}/(1-\hat{\delta}))\hat{\alpha}_2[1-\hat{\delta}(1+\hat{\beta}'\hat{\alpha})]^{-1} = 0.55$. The estimates of the variance ratio and noise ratio between the two series are close to those reported by Engsted, see Table 3. In this case A is given by (14) and it is seen that

$$f(A) = \frac{\delta}{1-\delta} A(I_l - \delta A)^{-1} = \frac{\delta}{(1-\delta)(1-\delta(1+\beta'\alpha))} A$$

so that with $a = (1,0,0)'$ and $b = (0,0,1)'$ we find $b'f(A) = \gamma a'$. Therefore

$$(a'\Sigma_Z a, a'\Sigma_Z f(A)'b, b'f(A)\Sigma_Z f(A)'b) = \Sigma_{Z11}(1, \gamma, \gamma^2)$$

which shows that $\rho = sign(\gamma), vr = \gamma^{-2}, nr = (1-\gamma)^2$ and that the estimates and their variance are functions of α and β, and their estimates and their variances. Hence the asymptotic distributions are identical, whether the covariance matrix of $Z_t' = (X_t'\beta, \Delta X_t')'$ is considered as fixed or not.

5 Conclusion

In Campbell and Shiller (1987) a method based on comparing actual and theoretical spreads was introduced. In this paper we have reconsidered the proposals in an I(1)

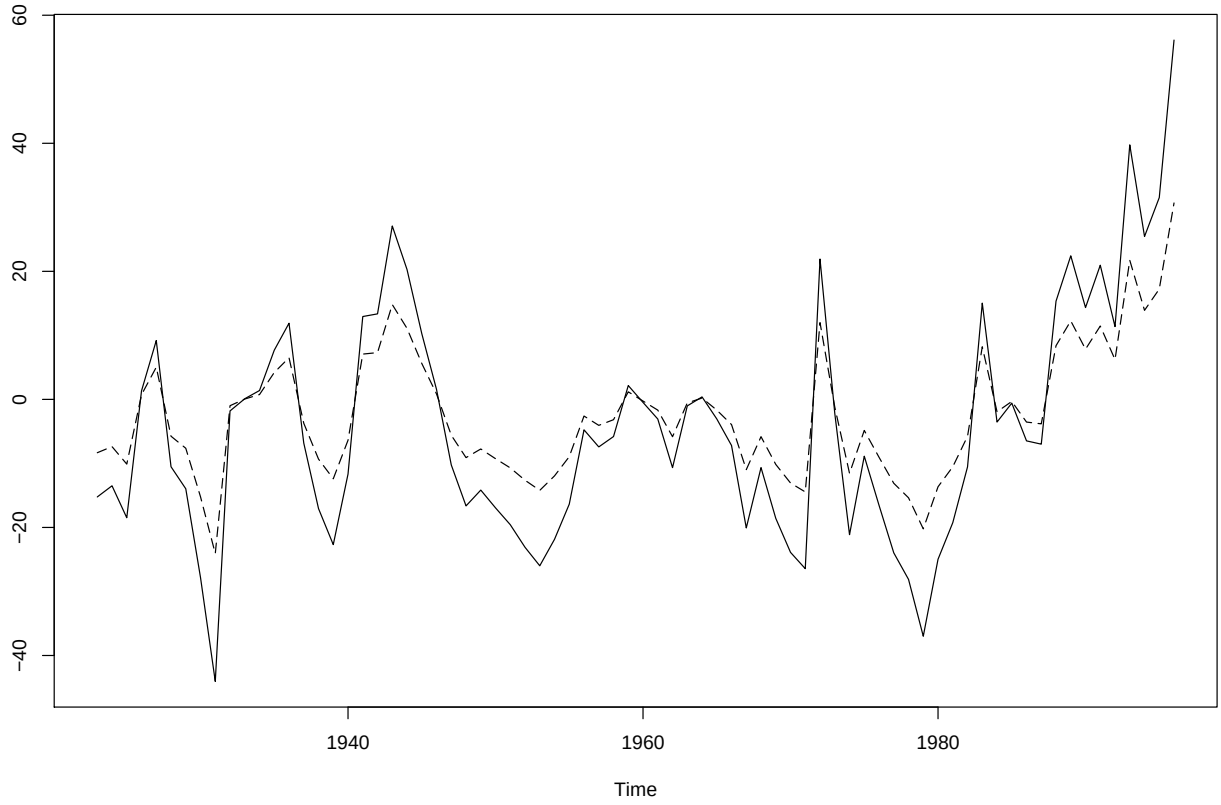

Figure 6: Plot of the estimates of actual spread (solid line) and first theoretical spread, S_t^2, (dashed line) of Danish stocks 1922-1996.

framework with a particular view to how the basic ideas can be extended to higher dimensions than the bivariate setup originally treated. We pointed out several problems that had to be resolved, proposed a statistical model for that purpose, and illustrated its potentiality in a number of examples. The asymptotic distribution of the difference of the spreads turns out to not exist as a process. However, the distributions of the commonly used numerical summaries, such as correlation, variance ratio and noise ratio, converge in distribution. The limiting distributions are Gaussian, and corresponding invariance principles have been worked out. Finally, we consider the additional insights that can be gained from applying the setup and methods to the data on U.S. stock prices and dividends originally studied by Campbell and Shiller (1987), and of the data on Danish stocks illustrated by Engsted (2002).

6 Appendix

We first give some general results in Lemma 1 and Lemma 2 about the asymptotic behavior of product moments derived from a multivariate AR(1) process. In Lemma 3 and Lemma 4 we then find the asymptotic properties of the estimators of the statistical model defined in Section 2.

Lemma 1 *i) Let X_i of dimension n_i, $i = 1, \ldots, 4$ be multivariate Gaussian random variables and let $Cov(X_i, X_j) = \Sigma_{ij}$. Then*

$$Cov(vec(X_1X_2'), vec(X_3X_4')) = (\Sigma_{24} \otimes \Sigma_{13}) + (\Sigma_{23} \otimes \Sigma_{14})K_{n_3n_4},$$

where $K_{n_3n_4}$ is the commutation matrix defined by $vec(M) = K_{n_3n_4}vec(M')$ for any $n_3 \times n_4$ matrix M, see for instance Lütkepohl (2005, p. 663).

ii) Let X_{it} of dimension n_i, $i = 1, \ldots, 4$ be multivariate Gaussian linear stationary processes with finite variances and let $Cov(X_{it}, X_{js}) = \gamma_{ij}(t-s)$. Then

$$\begin{aligned} &T^{-1}Cov(\sum_{t=1}^{T} vec(X_{1t}X_{2t}'), \sum_{t=1}^{T} vec(X_{3t}X_{4t}')) \\ &\rightarrow \sum_{h=-\infty}^{\infty} (\gamma_{24}(h) \otimes \gamma_{13}(h) + \gamma_{23}(h) \otimes \gamma_{14}(h)K_{n_3n_4}). \end{aligned}$$

Proof of i). The result is a multivariate version of the well known result for Gaussian univariate variables that

$$Cov(X_1X_2, X_3X_4) = Cov(X_1, X_3)Cov(X_2, X_4) + Cov(X_1, X_4)Cov(X_2, X_3).$$

It can be generalized to non Gaussian variables by adding an extra term depending on the fourth cumulant, see e.g. Anderson (1971, Thorem 8.4.2).

Let $B_{n_2 \times n_1}$ and $C_{n_4 \times n_3}$ be arbitrary matrices. A general linear function of $vec(X_1X_2')$ is $vec(B')'vec(X_1X_2') = tr\{BX_1X_2'\} = X_2'BX_1$ and we therefore investigate

$$\begin{aligned} Cov(X_2'BX_1, X_4'CX_3) &= \sum_{ijkm} B_{ij}C_{km}Cov(X_{2i}X_{1j}, X_{4k}X_{3m}) \\ &= \sum_{ijkm} B_{ij}C_{km}(\Sigma_{24,ik}\Sigma_{13,jm} + \Sigma_{23,im}\Sigma_{14,jk}) \\ &= tr\{B\Sigma_{13}C'\Sigma_{24}'\} + tr\{B\Sigma_{14}C\Sigma_{23}'\} \\ &= vec(B')'(\Sigma_{24} \otimes \Sigma_{13})vec(C') + vec(B')'(\Sigma_{23} \otimes \Sigma_{14})vec(C) \\ &= vec(B')'[(\Sigma_{24} \otimes \Sigma_{13}) + (\Sigma_{23} \otimes \Sigma_{14})K_{n_3n_4}]vec(C'), \end{aligned}$$

where we apply the formula $tr(ABCD) = vec(A')(D' \otimes B)vec(C)$.
Proof of ii). The covariance equals

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} Cov(vec(X_{1t}X_{2t}'), vec(X_{3s}X_{4s}')) \\
&= T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}[(\gamma_{24}(t-s)\otimes\gamma_{13}(t-s)) + (\gamma_{23}(t-s)\otimes\gamma_{14}(t-s))K_{n_3n_4}] \\
&= T^{-1}\sum_{h=-T}^{T}(T-|h|)[(\gamma_{24}(h)\otimes\gamma_{13}(h)) + (\gamma_{23}(h)\otimes\gamma_{14}(h))K_{n_3n_4}]
\end{aligned}$$

which converges as indicated.

Lemma 2 *Let $Y_t = AY_{t-1} + \delta_t$ be an $l-$dimensional stationary vector autoregressive process and let δ_t be i.i.d. $N_l(0,\Phi)$. Then*

(i) $Y_t = \sum_{i=0}^{\infty} A^i\delta_{t-i}$ *has mean zero and variance* $\Sigma_Y = \sum_{i=0}^{\infty} A^i\Phi A^{i\prime}$ *which can be written* $vec(\Sigma_Y) = (I_{l^2} - A\otimes A)^{-1}vec(\Phi)$.

(ii) $T^{-1}\sum_{t=1}^{T} Y_t \xrightarrow{P} 0$ *and* $T^{-1}\sum_{t=1}^{T} Y_{t-1}Y_{t-1}' \xrightarrow{P} \Sigma_Y$.

(iii) Let

$$V_{1,T} = \sum_{t=1}^{T} Y_{t-1}\delta_t' \text{ and } V_{2,T} = \sum_{t=1}^{T}(Y_{t-1}Y_{t-1}' - \Sigma_Y)$$

and $V_T = (vec(V_{1,T}')', vec(V_{2,T}')')'$, *then*

$$\lim_{T\to\infty} T^{-1}Var(V_T) = \Sigma_V = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{27}$$

where

$$\begin{aligned}
\Sigma_{11} &= \Sigma_Y \otimes \Phi \\
\Sigma_{21} &= (I_{l^2} - A\otimes A)^{-1}[(A\Sigma_Y\otimes\Phi) + (\Phi\otimes A\Sigma_Y)K_{ll}] \\
\Sigma_{22} &= [(\Sigma_Y\otimes\Sigma_Y + (I_{l^2} - A\otimes A)^{-1}(A\Sigma_Y\otimes A\Sigma_Y) \\
&\quad + (\Sigma_Y A'\otimes\Sigma_Y A')(I_{l^2} - A'\otimes A')^{-1})](I_{l^2} + K_{ll})
\end{aligned}$$

(iv) Finally

$$T^{-1/2}V_{[Tu]} \Longrightarrow W_V(u)$$

as processes on $D^{(2l)^2}[0,1]$ where $W_V(u)$ is an $(2l)^2-$dimensional Brownian motion with variance Σ_V.

Proof of (i). We find

$$Y_t = AY_{t-1} + \delta_t = \sum_{i=0}^{\infty} A^i \delta_{t-i},$$

which has mean zero and variance

$$\sum_{i=0}^{\infty} A^i \Phi A^{i\prime},$$

which in vectorized form is $(I_{l^2} - A \otimes A)^{-1} vec(\Phi)$.

Proof of (ii). The convergence follows from the law of large numbers.

Proof of (iii). We define

$$\begin{aligned}
&\gamma(h) = Cov(Y_{t+h}, Y_t) = A^h \Sigma_Y \text{ if } h \geq 0 \text{ and } \Sigma_Y A'^{-h} \text{ if } h < 0,\\
&\phi(h) = Cov(Y_{t+h}, \delta_t) = A^h \Phi \text{ if } h \geq 0 \text{ and } 0 \text{ if } h < 0,\\
&\eta(h) = Cov(\delta_{t+h}, \delta_t) = \Phi \text{ if } h = 0 \text{ and } 0 \text{ if } h \neq 0.
\end{aligned}$$

From Lemma 1 we find that

$$\begin{aligned}
&T^{-1} Var(vec(V'_{2,[Tu]}))\\
&\rightarrow \sum_{h=-\infty}^{\infty} [(\gamma(h) \otimes \gamma(h)) + (\gamma(h) \otimes \gamma(h)) K_{ll}]\\
&= (\Sigma_Y \otimes \Sigma_Y + \sum_{h=1}^{\infty} (A^h \Sigma_Y \otimes A^h \Sigma_Y) + \sum_{h=1}^{\infty} (\Sigma_Y A^{h\prime} \otimes \Sigma_Y A^{h\prime}))(I_{l^2} + K_{ll})\\
&= [\Sigma_Y \otimes \Sigma_Y + (I_{l^2} - A \otimes A)^{-1} (A\Sigma_Y \otimes A\Sigma_Y)\\
&+ (\Sigma_Y A' \otimes \Sigma_Y A')(I_{l^2} - A' \otimes A')^{-1}](I_{l^2} + K_{ll}).
\end{aligned}$$

Similarly

$$\begin{aligned}
&T^{-1} Cov(vec(V'_{2,[Tu]}), vec(V'_{1,[Tu]}))\\
&= T^{-1} Cov(\sum_{t=1}^{T} vec(Y_{t-1} Y'_{t-1}), \sum_{t=1}^{T} vec(\delta_t Y'_{t-1}))\\
&\rightarrow \sum_{h=-\infty}^{\infty} [(\gamma(h) \otimes \phi(h-1)) + (\phi(h-1) \otimes \gamma(h)) K_{ll}]\\
&= \sum_{h=1}^{\infty} [(A^h \Sigma_Y \otimes A^{h-1} \Phi) + (A^{h-1} \Phi \otimes A^h \Sigma_Y) K_{ll}]\\
&= (I_{l^2} - A \otimes A)^{-1} [(A\Sigma_Y \otimes \Phi) + (\Phi \otimes A\Sigma_Y) K_{ll}].
\end{aligned}$$

Finally

$$Var(vec(V'_{1,T})) = Var(T^{-1/2}\sum_{t=1}^{T} vec(\delta_t Y'_{t-1}))$$
$$= \sum_{h=-\infty}^{\infty} [\gamma(h) \otimes \eta(h)) + (\phi(h-1) \otimes \phi(-h+1)')K_{ll}] = \Sigma_Y \otimes \Phi.$$

Proof of (iv). We can apply Theorem 21.1 in Billingsley (1968) which gives an invariance principle for $\phi-$mixing sequences. We study the joint asymptotic distribution of $\sum_{t=1}^{T} Y_{t-1}Y'_{t-1}, \sum_{t=1}^{T} Y_{t-1}\delta'_t$ by considering the univariate process

$$\eta_t = f(\delta_t, \delta_{t-1}, \dots) = tr\{B(Y_t Y'_t - \Sigma_Y)\} + tr\{C(\delta_{t+1}Y'_t)\},$$

for arbitrary $l \times l$ matrices B and C. We define an $m-$ dependent approximation to Y_t by

$$Y_t^m = \sum_{i=0}^{m} A^i \delta_{t-i}$$

with the remainder term

$$Y_t^r = \sum_{i=m+1}^{\infty} A^i \delta_{t-i} = A^m \sum_{i=1}^{\infty} A^i \delta_{t-m-i},$$

so that

$$\Sigma_Y = \Sigma_Y^m + \Sigma_Y^r = Var(Y_t^m) + Var(Y_t^r).$$

We define

$$\eta_t^m = tr\{B(Y_t^m Y_t^{m\prime} - \Sigma_Y^m)\} + tr\{C(\delta_{t+1}Y_t^{m\prime})\}$$

and have to show that for

$$\nu_m = E(\eta_t - \eta_t^m)^2$$

it holds that

$$\sum_{m=1}^{\infty} \nu_m^{1/2} < \infty.$$

This follows if we can show that ν_m is exponentially decreasing in m.

We find using $Y_t = Y_t^m + Y_t^r$ that

$$\eta_t - \eta_t^m = tr\{B(Y_t^m Y_t^{r\prime} + Y_t^r Y_t^{m\prime} + Y_t^r Y_t^{r\prime} - \Sigma_Y^r)\} + tr\{C\delta_{t+1}Y_t^{r\prime}\},$$

so that all terms contain an exponentially decreasing factor A^m, which implies that $\sum_{m=1}^{\infty} \nu_m^{1/2} < \infty$ and Theorem 21.1 of Billingsley (1968) now gives the required result. The variance was found in (iii). ■

Lemma 3 *Let the $I(1)$ process X_t be given by the cointegrated VAR model with restricted linear term, (8). Assume that β is identified by the restrictions*

$$\beta_i^* = h_i^* + H_i^* v_i, i = 1, \ldots, r,$$

where h_i^* is a $(p+1)\times 1$ vector and H_i^* is $(p+1)\times m_i$ so that β_i^* has m_i free parameters. Let ξ denote the trend coefficient in the process X_t, see (9), then $H_i^{*\prime} X_t^*$ has trend coefficient

$$\tau_i = H_i^{*\prime} \begin{pmatrix} \xi \\ 1 \end{pmatrix}.$$

If $\tau_i \neq 0$, we define the normalization matrix

$$A_{iT} = \left(T^{-1/2}\tau_{i\perp}, T^{-1}\bar{\tau}_i\right)$$

where $\bar{\tau}_i = \tau_i(\tau_i'\tau_i)^{-1}$, and let $H_i = (I_p, 0)H_i^*$. Then

$$A_{iT}' H_i^{*\prime} X_{[Tu]}^* \Longrightarrow \begin{pmatrix} \tau_{i\perp}' H_i' CW(u) \\ u \end{pmatrix} = G_i(u)$$

and the limit in distribution of $\{T^{1/2}A_{iT}^{-1}(\hat{\beta}_i^* - \beta_i^*)\}$ is the mixed Gaussian distribution

$$\{B_i\} = \{\alpha_i'\Omega^{-1}\alpha_j \int_0^1 (G_i - \bar{G}_i)(G_i - \bar{G}_i)'du\}^{-1}\{\int_0^1 (G_i - \bar{G}_i)(dW)'\Omega^{-1}\alpha_i\},$$

where $\bar{G}_i = \int_0^1 G_i(u)du$. Finally we have the results

$$T^{1/2}(\hat{\beta}_i^* - \beta_i^*)' X_{[Tu]}^* \Longrightarrow B_i' G_i(u) \tag{28}$$

$$\sum_{t=1}^{T} (\hat{\beta}_i^* - \beta_i^*)' X_t^* X_t^{*\prime} (\hat{\beta}_i^* - \beta_i^*) \Longrightarrow B_i' \int_0^1 G_i(u) G_i(u)' du B_i \tag{29}$$

A similar result holds if $\tau_i = 0$, but then $G_i = H_i CW(u)$.

It follows that likelihood ratio tests for hypotheses on β^* are asymptotically distributed as χ^2.

Proof: From (9) it follows that

$$H_i^{*\prime} X_{[Tu]}^* = H_i^{*\prime} \begin{pmatrix} \xi[Tu] \\ [Tu]+1 \end{pmatrix} + H_i^{*\prime} \begin{pmatrix} C\sum_{i=1}^{[Tu]} \varepsilon_t \\ 0 \end{pmatrix} + H_i^{*\prime} \begin{pmatrix} A_0 + \xi_0 + Y_{[Tu]} \\ 0 \end{pmatrix}. \tag{30}$$

so that in the direction τ_i, the linear term dominates, and orthogonal to this the random walk dominates:

$$T^{-1}\bar{\tau}_i' H_i^{*\prime} X_{[Tu]}^* \Longrightarrow u$$

$$T^{-1/2}\tau_{i\perp}' H_i^{*\prime} X_{[Tu]}^* \Longrightarrow \tau_{i\perp}' H_i^{*\prime} \begin{pmatrix} CW(u) \\ 0 \end{pmatrix} = \tau_{i\perp}' H_i' CW(u),$$

which we write as

$$A'_{iT}H_i^{*\prime}X^*_{[Tu]} \Longrightarrow G_i(u).$$

Let R_{1t} denote the residuals of X^*_{t-1} corrected for lagged differences and a constant. Then the likelihood profile for the parameters (α, β, Ω) gives the score for υ_i normalized by A_{iT}, which converges in distribution to a mixed Gaussian distribution

$$\sum_{t=1}^{T} A'_{iT}H_i^{*\prime}R_{1t}\varepsilon'_t\Omega^{-1}\alpha_i \Longrightarrow \int_0^1 (G_i - \bar{G}_i)(dW)'\Omega^{-1}\alpha_i$$

and the information for (υ_i, υ_j) normalized by $T^{-1/2}A'_{iT}$ and $T^{-1/2}A_{jT}$ converges in distribution

$$\alpha'_i\Omega^{-1}\alpha_j T^{-1}\sum_{t=1}^{T} A'_{iT}H_i^{*\prime}R_{1t}R'_{1t}H_j^*A_{jT} \Longrightarrow \alpha'_i\Omega^{-1}\alpha_j \int_0^1 G_i(u)G'_j(u)du,$$

so that the asymptotic distribution of $\{T^{1/2}A_{iT}^{-1}(\hat{\beta}_i^* - \beta_i^*)\}$ is mixed Gaussian as indicated.

The asymptotic conditional variance matrix is non-singular because identification implies that $\beta_\perp^{*\prime}H_i^*$ has full rank m_i, see Johansen (2009, Lemma 4). An estimator of the asymptotic conditional variance is given by

$$\{\hat{A}'_{iT}H_i^{*\prime}S_{11}^*H_j^{*\prime}\hat{A}_{jT}\}, \tag{31}$$

where $TS_{11}^* = \sum_{t=1}^{T} R_{1t}R'_{1t}$. ■

Lemma 4 *Let the I(1) process X_t be given by (8) and let Σ_Z be the asymptotic variance of the stacked process of dimension $l = r + (k-1)p$. Then, as $T \to \infty$, $T^{1/2}vec(\hat{A} - A)$ converges in distribution to $N_{l^2}(0, \Sigma_Z^{-1} \otimes Q\Omega Q')$ when $k > 1$. For $k = 1$, the limit is $N_{(r+p)^2}(0, (I_r, 0)'((I_r, 0)\Sigma_Z(I_r, 0)')^{-1}(I_r, 0) \otimes Q\Omega Q')$.*

Proof. When $k > 1$

$$T^{1/2}(\hat{A} - A) = QT^{1/2}\left(\hat{\alpha} - \alpha, \hat{\Gamma}_1 - \Gamma_1, \ldots, \hat{\Gamma}_{k-1} - \Gamma_{k-1}\right) + o_P(1) = T^{1/2}Q(\hat{\theta} - \theta) + o_P(1).$$

The rest of the proof is as in Theorem 13.5 in Johansen (1995). The rows $r+1, \ldots, r+p$ of $Z_t = AZ_{t-1} + \underline{\mu} + Q\varepsilon_t$ can be written

$$\Delta X_t = \theta Z_{t-1}(\beta^*) + \mu_0 + \varepsilon_t.$$

For fixed β^* this is a regression equation and

$$\begin{aligned} T^{1/2}Q(\hat{\theta} - \theta) &= T^{-1/2}\sum_{i=1}^{T} Q\varepsilon_t(Z_{t-1}(\beta^*) - \mu_Z)'\Sigma_Z^{-1} + o_P(1) \\ &= T^{-1/2}V'_{1,T}\Sigma_Z^{-1} + o_P(1), \end{aligned}$$

because, referring to Lemma 2, $\delta_t = Q\varepsilon_t$ and $Y_{t-1} = Z_{t-1}(\beta^*) - \mu_Z$. Since $\Phi = Q\Omega Q'$, it also follows from Lemma 2 that $T^{1/2}vec(Q(\hat{\theta} - \theta))$ converges in distribution toward $N_{l^2}(0, \Sigma_Z^{-1} \otimes Q\Omega Q')$.

When $k = 1$

$$\sqrt{T}(\hat{A} - A) = \begin{pmatrix} \hat{\beta}'\hat{\alpha} - \beta'\alpha & 0 \\ \hat{\alpha} - \alpha & 0 \end{pmatrix} = Q\sqrt{T}(\hat{\alpha} - \alpha, 0) + o_P(1) = \sqrt{T}Q(\hat{\alpha} - \alpha)(I_r\ 0) + o_P(1),$$

which shows the result.■

Proof of Proposition 1. We first show that replacing $\hat{Z}_t$ with Z_t in $\bar{S}_T$ we get a difference of $O_P(T^{-1/2})$. From (28) and (29) in Lemma 3 we have that the difference $\hat{Z}_t - Z_t$ is $O_P(T^{-1/2})$ because

$$T^{1/2}(\hat{Z}_{[Tu]} - Z_{[Tu]})' = (X'_{[Tu]}(\hat{\beta}_1 - \beta_1), 0, \ldots, 0) \Longrightarrow (G'_1(u)B_1, 0, \ldots, 0). \tag{32}$$

Then, see (20),

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} S_t - (\psi T^{-1}\sum_{t=1}^{T}\hat{Z}_t + \eta) &= \psi T^{-1}\sum_{t=1}^{T} Z_t + \eta - (\psi T^{-1}\sum_{t=1}^{T}\hat{Z}_t + \eta) \\ &= \psi T^{-1}\sum_{t=1}^{T}(Z_t - \hat{Z}_t) = O_P(T^{-1/2}). \end{aligned} \tag{33}$$

Next we show that replacing $\hat{Z}_t$ with Z_t and replacing $\hat{\mu}_S$ by μ_S (or $\hat{\bar{Z}}_T$ by μ_Z) in U_T gives a difference of $O_P(T^{-1})$. Because $Y_t = Z_t - \mu_Z$, we may write $\hat{Z}_t - \hat{\mu}_Z = (\hat{Z}_t - Z_t) + Y_t + (\mu_Z - \hat{\mu}_Z)$ and find

$$\begin{aligned} &T^{-1}[\sum_{t=1}^{[Tu]}(\hat{Z}_t - \hat{\mu}_Z)(\hat{Z}_t - \hat{\mu}_Z)' - \sum_{t=1}^{[Tu]} Y_t Y_t'] \\ &= T^{-1}\sum_{t=1}^{[Tu]}(\hat{Z}_t - Z_t)(\hat{Z}_t - Z_t)' + T^{-1}\sum_{t=1}^{[Tu]} Y_t(\hat{Z}_t - Z_t)' + T^{-1}\sum_{t=1}^{[Tu]}(\hat{Z}_t - Z_t)Y_t' \\ &+ T^{-1}[Tu](\mu_Z - \hat{\mu}_Z)(\mu_Z - \hat{\mu}_Z)' + [T^{-1}\sum_{t=1}^{[Tu]} Y_t](\hat{\mu}_Z - \mu_Z)' + (\hat{\mu}_Z - \mu_Z)[T^{-1}\sum_{t=1}^{[Tu]} Y_t'] \\ &+ [T^{-1}\sum_{t=1}^{[Tu]}(\hat{Z}_t - Z_t)](\hat{\mu}_Z - \mu_Z)' + (\hat{\mu}_Z - \mu_Z)[T^{-1}\sum_{t=1}^{[Tu]}(\hat{Z}_t - Z_t)']. \end{aligned} \tag{34}$$

Now from (29) of Lemma 3 we find that the first term is $O_P(T^{-1})$. The second and third term are

$$T^{-1}\sum_{t=1}^{[Tu]}(\hat{Z}_t - Z_t)Y_t' = (I_r, 0, \ldots, 0)'T^{-1}\sum_{t=1}^{[Tu]}(\hat{\beta}_1^* - \beta_1^*)'X_t^*Y_t'$$
$$= T^{-1}(I_r, 0, \ldots, 0)'T^{-1/2}\sum_{t=1}^{[Tu]}T^{1/2}[A_{T1}^{-1}(\hat{\beta}_1^* - \beta_1^*)]'A_{T1}'X_t^*Y_t' = O_P(T^{-1}).$$

The next three terms are $O_P(T^{-1})$ because from from Lemma 2 with $Y_t = Z_t - \mu_Z$, $\delta_t = Q\varepsilon_t$, and $\Phi = Q\Omega Q'$, $\hat{\mu}_Z - \mu_Z = O_P(T^{-1/2})$ and $T^{-1}\sum_{t=1}^{[Tu]}Y_t = O_P(T^{-1/2})$. Finally the last two terms are products of terms which we have argued each is $O_P(T^{-1/2})$.

We can now prove the relations (22) and (23). The law of large numbers shows that

$$\bar{S}_T \xrightarrow{P} \psi\mu_Z + \eta = \mu_S,$$
$$U_T \xrightarrow{P} \psi\Sigma_Z\psi' = \Sigma_S,$$

and we have shown above that we can replace $\hat{Z}_t$ by Z_t and μ_S by its estimate with a small error, and Lemma 4 shows we can replace (ψ, η) by their estimates, so that (22) and (23) follows.

We can now find the asymptotic distribution of

$$uT^{1/2}(\hat{U}_{[Tu]} - \psi\Sigma_Z\psi') = T^{-1/2}\sum_{t=1}^{[Tu]}[\hat{\psi}_{[Tu]}(\hat{Z}_t - \hat{\mu}_Z)(\hat{Z}_t - \hat{\mu}_Z)'\hat{\psi}_{[Tu]}' - \psi\Sigma_Z\psi']. \tag{35}$$

From (34) it follows that the difference between U_T and what we get by inserting $\hat{Z}_t$ for Z_t and $\hat{\mu}_Z$ for μ_Z is $O_P(T^{-1})$.

Thus for $Y_t = Z_t - \mu_Z$, (35) has the same limit distribution as

$$T^{-1/2}\sum_{t=1}^{[Tu]}[\hat{\psi}_{[Tu]}Y_tY_t'\hat{\psi}_{[Tu]}' - \psi\Sigma_Z\psi'] \tag{36}$$
$$= T^{1/2}\psi\sum_{t=1}^{[Tu]}(Y_tY_t' - \Sigma_Z)\psi' + T^{1/2}(\hat{\psi}_{[Tu]} - \psi)u\Sigma_Z\psi' + T^{1/2}\psi u\Sigma_Z(\hat{\psi}_{[Tu]} - \psi)' + o_P(1).$$

With the notation $V_{1,T}$ and $V_{2,T}$ from Lemma 2, using that $\delta_t = Q\varepsilon_t$, and $\zeta = (0, b)'$ we find

$$
\begin{aligned}
uT^{1/2}vec((\hat{\psi}_{[Tu]} - \psi)\Sigma_Z\psi') &= uT^{1/2}vec(\zeta(f(\hat{A}_{[Tu]}) - f(A))\Sigma_Z\psi') \\
&= uT^{1/2}(\psi\Sigma_Z \otimes \zeta)vec(f(\hat{A}_{[Tu]}) - f(A)) \\
&= uT^{1/2}(\psi\Sigma_Z \otimes \zeta)Fvec(\hat{A}_{[Tu]} - A) + o_P(1) \\
&= (\psi\Sigma_Z \otimes \zeta)Fvec(T^{-1/2}V'_{1,[Tu]}\Sigma_Z^{-1}) + o_P(1) \\
&= (\psi\Sigma_Z \otimes \zeta)F(\Sigma_Z^{-1} \otimes I_l)vec(T^{-1/2}V'_{1,[Tu]}) + o_P(1).
\end{aligned}
$$

Therefore we can write (36) in vectorized form as

$$
\begin{aligned}
&(\psi \otimes \psi)vec(V'_{2,[Tu]}) \\
&+ [(\psi\Sigma_Z \otimes \zeta) + (\zeta \otimes \psi\Sigma_Z)K_{ll}]F(\Sigma_Z^{-1} \otimes I_l)]vec(V'_{1,[Tu]}) + o_P(T^{1/2}).
\end{aligned} \tag{37}
$$

When $k = 1$, we use $(I_r\ 0)'((I_r\ 0)\Sigma_Z(I_r\ 0)')^{-1}(I_r\ 0)$ instead of Σ_Z^{-1}.

From Lemma 2 with $Y_t = Z_t - \mu_Z$, $\delta_t = Q\varepsilon_t$, and $\Phi = Q\Omega Q'$ we find that $uT^{1/2}vec(\hat{U}_{[Tu]} - \Sigma_S)$ converges in distribution on $D^4[0,1]$ to

$$
[(\zeta \otimes \psi\Sigma_Z)K_{ll} + (\psi\Sigma_Z \otimes \zeta)]F(\Sigma_Z^{-1} \otimes I_l), \psi \otimes \psi)V,
$$

which has a variance as indicated.

Finally we find, replacing Σ_S by $\hat{U}_T$, that

$$
uT^{1/2}vec(\hat{U}_{[Tu]} - \hat{U}_T) \Longrightarrow W_U(u) - uW_U(1). \quad \blacksquare
$$

References

[1] Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] Campbell, J.Y. and R.J. Shiller (1987). Cointegration tests and present value models. *Journal of Political Economy* **95**, 1062-1088.

[4] Campbell, J.Y. and R.J. Shiller (1989). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies* **1**, 195-228.

[5] Carriero, A., Favero, C.A. and I. Kaminska (2006). Financial factors, macroeconomic information and the Expectations Theory of the term structure of interest rates. *Journal of Econometrics* 131, 339-358.

[6] Engsted, T. (2002). Measures of fit for rational expectations models. *Journal of Economic Surveys* **16**, 301-355.

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

[8] Johansen, S. (1995b). Identifying Restrictions of Linear Equations — with applications to simultaneous equations and cointegration. *Journal of Econometrics* 69, 111–132.

[9] Johansen, S. (2009). Some identifying problems in the cointegrated vector autoregressive model, forthcoming *Journal of Econometrics*.

[10] Johansen, S. and A.R. Swensen (1999). Testing Rational Expectations in Vector Autoregressive Models. *Journal of Econometrics* 93, 73-91.

[11] Johansen, S. and A.R. Swensen (2004). More on testing exact rational expectations in vector autoregressive models: Restricted drift term, *The Econometric Journal,* 7, 389-397 (2004).

[12] Kurmann, A. (2005). Quantifying the uncertainty about the fit of a new Keynesian pricing model. *Journal of Monetary Economics* 52, 1119-1134.

[13] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, New York.

Research Papers 2009

CREATES
Center for Research in Econometric Analysis of Time Series

2009-5: Tim Bollerslev, Natalia Sizova and George Tauchen: Volatility in Equilibrium: Asymmetries and Dynamic Dependencies

2009-6: Anders Tolver Jensen and Theis Lange: On IGARCH and convergence of the QMLE for misspecified GARCH models

2009-7: Jeroen V.K. Rombouts and Lars Stentoft: Bayesian Option Pricing Using Mixed Normal Heteroskedasticity Models

2009-8: Torben B. Rasmussen: Jump Testing and the Speed of Market Adjustment

2009-9: Dennis Kristensen and Andrew Ang: Testing Conditional Factor Models

2009-10: José Fajardo and Ernesto Mordecki: Skewness Premium with Lévy Processes

2009-11: Lasse Bork: Estimating US Monetary Policy Shocks Using a Factor-Augmented Vector Autoregression: An EM Algorithm Approach

2009-12: Konstantinos Fokianos, Anders Rahbek and Dag Tjøstheim: Poisson Autoregression

2009-13: Peter Reinhard Hansen and Guillaume Horel: Quadratic Variation by Markov Chains

2009-14: Dennis Kristensen and Antonio Mele: Adding and Subtracting Black-Scholes: A New Approach to Approximating Derivative Prices in Continuous Time Models

2009-15: Charlotte Christiansen, Angelo Ranaldo and Paul Söderllind: The Time-Varying Systematic Risk of Carry Trade Strategies

2009-16: Ingmar Nolte and Valeri Voev: Least Squares Inference on Integrated Volatility and the Relationship between Efficient Prices and Noise

2009-17: Tom Engsted: Statistical vs. Economic Significance in Economics and Econometrics: Further comments on McCloskey & Ziliak

2009-18: Anders Bredahl Kock: Forecasting with Universal Approximators and a Learning Algorithm

2009-19: Søren Johansen and Anders Rygh Swensen: On a numerical and graphical technique for evaluating some models involving rational expectations