Chapter 6: Multivariate financial time series analysis

Contents

1. **Stochastic trends and spurious regressions**

From what we have discussed so far, it should be clear that econometric analysis depends on the variances and covariances among variables. In the case of independent sampling (like in cross-sectional applications) over time, we can use finite sample moments for estimation and inference, and in the case of stationary time series the consideration of moments in large samples can solve the problems peculiar to time series in small samples. Within this framework non-stationarity causes problems: we know that unconditional moments are not defined for non-stationary time series. Consider, for the sake of illustration, an OLS regression of an I(0) variable y_t on an I(1) variable x_t. The OLS estimator of the regression y_t on x_t converges to zero as the sample size increases. The variance of x_t, being divergent, dominates the covariance between the two variables. In general, asymptotic theory is not applicable to non-stationary time-series (see, for example, Hatanaka, 1996; Maddala and Kim, 1998). So, unless all the trends observed in time series are deterministic, the solution of reverting to asymptotic theory is not directly accessible.

To give some intuition on the importance of non-stationarity in time series analysis and to illustrate the problems related to non-stationarity, consider the results of two regressions reported in Table 1, obtained by regressing the (natural) logarithm of UK stock prices on the log of US dividends and the log of UK dividends.

Table 1. Regressing UK log-prices on US log-dividends and UK log-dividends

SAMPLE 1973:1-2011:4, Dep. Var LPUK			
Variable	Coefficient	Std. Error	t-Statistic
C	3.38	0.086	38.18
LDUS	1.71	0.036	44.46
$R^2 = 0.9295$	S.E. of regression: 0.0853	DW statistic: 0.13	
C	2.31	0.069	33.04
LDUK	1.22	0.017	72.07
$R^2 = 0.972$	S.E. of regression: 0.033	DW statistic: 0.32	

This regression features an extremely high R^2 and significant coefficients and it apparently difficult to distinguish empirically between the two models. In fact, the regression of UK prices on US dividends is a case of a spurious regression, which witnesses the relevance of non-stationarity in economic time series. To elaborate on this point, consider the two simple univariate time-series models for LPUK and LDUS shown in Table 2.

TABLE 2. Univariate models for UK prices and US dividends

SAMPLE 1973:1-2011:4			
Variable	Coefficient	Std. Error	t-Statistic
Dependent variable LPUK			
C	0.111	0.153	0.728
LPUK(-1)	0.986	0.023	25.49
$R^2 = 0.99$ S.E. of regression: 0.042			
Dependent variable LDUS			
C	0.0214	0.261	0.082
LDUS(-1)	0.993	0.091	10.90
S.E. of regression: 0.011			

Despite the simplicity of the two models for LDUS and LPUK, we note that they can both be approximated by random walk models:

$$\begin{gathered}\text{LDUS}_t = a_0 + \text{LDUS}_{t-1} + \epsilon_{1t}, \\ \text{LPUK}_t = b_0 + \text{LPUK}_{t-1} + \epsilon_{2t}, \\ \epsilon_{1t} \sim NID\left(0, \sigma^2_{\epsilon_1}\right) \qquad \epsilon_{2t} \sim NID\left(0, \sigma^2_{\epsilon_2}\right).\end{gathered}$$

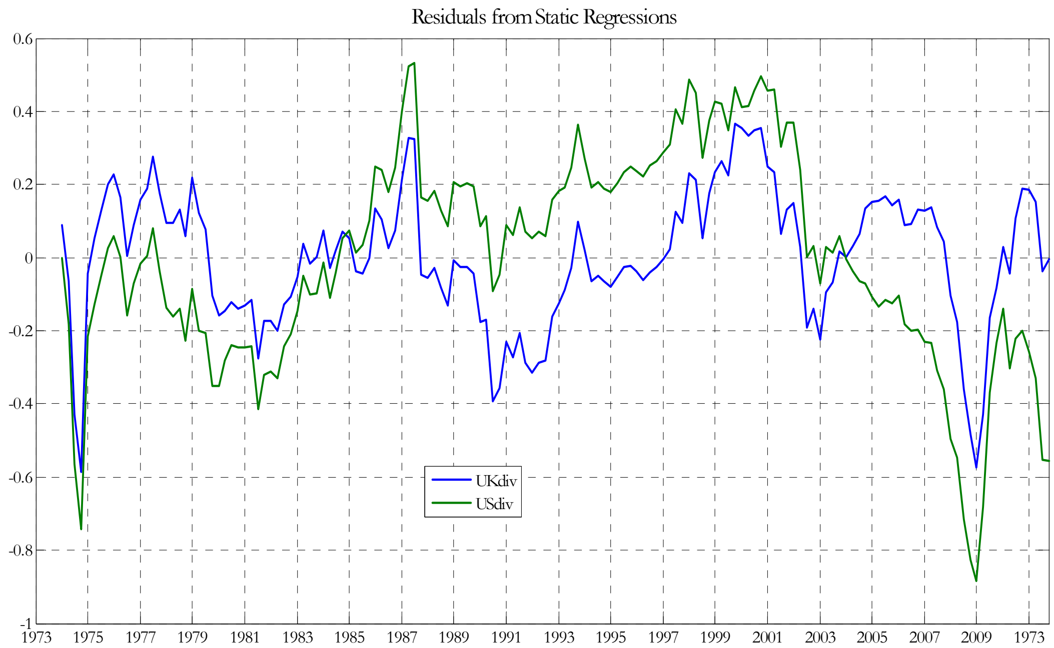
As we already know, recursive substitution yields:

$$\begin{aligned}\text{LDUS}_t &= \text{LDUS}_0 + a_0 t + \sum_{i=0}^{t-1} \epsilon_{1t-i}, \\ \text{LPUK}_t &= \text{LPUK}_0 + b_0 t + \sum_{i=0}^{t-1} \epsilon_{2t-i}.\end{aligned}$$

When the following model is estimated:

$$\text{LPUK}_t = \widehat{\alpha} + \widehat{\beta}\text{LDUS}_t + \widehat{\epsilon}_t,$$

the coefficient $\widehat{\beta}$ may be significant simply because both series display a deterministic trend. In a way, one is partly regressing $a_0 t$ on $b_0 t$ and, this being dominated by the deterministic trend, leads to a rather accurate fit. However, to have a non-spurious relation, we require that the regression also removes the stochastic trend from the dependent variables, leaving stationary residuals. Otherwise, the correlation we observe can be labelled as *spurious*. We report in the Figure below the residuals from the OLS regression of LPUK on LDUS and of LPUK on LDUK.



Visual impressions confirm the intuition that the regression has delivered a spurious relation, failing to remove the stochastic trend from the non-stationary dependent variable. The reported Durbin-Watson statistic of 0.13 gives a more formal background to the visual impression. The Durbin-Watson statistic, originally designed to test for the presence of first-order autocorrelation in the residuals, can be re-calibrated to test for stationarity:

$$\mathrm{DW} = \frac{\sum_{i=2}^{T} \left(\widehat{\epsilon}_t - \widehat{\epsilon}_{t-1}\right)^2}{\sum_{i=2}^{T} \widehat{\epsilon}_t^2} \simeq 2\left(1 - \widehat{\rho}\right),$$

where $\widehat{\rho}$ is the OLS coefficient from the regression of $\widehat{\epsilon}_t$ on $\widehat{\epsilon}_{t-1}$. The test was originally tabulated to test the hypothesis H_0: $\rho = 0$; however, critical values for the null of non-stationarity, H_0: $\rho = 1$, have been provided by Sargan and Bhargava (1983). According to such critical values, the null of non-stationarity cannot be rejected by an observed value of 0.14 for the Durbin-Watson statistic.

From an econometric point of view, non-stationarity of time series is problematic in that it might generate a spurious regression and it does not allow the use of standard large-sample theory for valid estimation and inference in the linear model. From an economic perspective, the presence of a unit root in the residuals means that the effects of a shock persist forever and that cyclical fluctuations cannot be studied separately from long-run growth components, as long-run trends are not fixed. Pioneering work by Nelson and Plosser (1982) has renewed the attention of the profession for these issues and after their work many tests have been proposed to discriminate between stochastic and deterministic trends. The Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) tests (1981) have enjoyed a remarkable empirical success. These are tests for the null hypothesis of non-stationarity of a generic time-series x_t, based on the following auxiliary regression:

$$x_t = \widehat{\mu} + \widehat{\gamma} t + \widehat{\delta} x_{t-1} + \sum_{i=0}^{k} \widehat{\varphi}_i \Delta x_{t-i-1} + \widehat{\epsilon}_t \tag{1}$$

under the unit root null hypothesis $\delta = 1$, therefore the test statistic is simply a t test: $t \equiv (\hat{\delta} - 1)/SE(\hat{\delta})$, where $SE(\hat{\delta})$ is the standard error of the estimated coefficient. The ADF statistic is obtained by selecting an appropriate value for k in (1), while the DF statistic is obtained by setting $k = 0$. Note that this statistic does not have the usual Student-t distribution, but it is skewed toward negative values. Dickey and Fuller (1981) have calculated the appropriate asymptotic critical values, which are affected by the specification of the deterministic component of (1). The ADF tests offer a solution to the dynamic mis-specification of the simple regressions behind DF and CRDW tests; alternative solutions generate different tests for the same null hypothesis (Phillips-Perron, 1988). It has been proved that the power of all these tests for a unit root against plausible deterministic trends alternative hypotheses is very limited (DeJong and Whiteman 1991). Rudebusch (1993) analyse the case of US real GNP and conclude that

> (...) The appropriate conclusion from unit-root test on this data sample is that the existence of a unit-root is uncertain (...)

The low power problem becomes even more complicated when deterministic trends with structural breaks are considered, as in Zivot and Andrews (1992). Interestingly, the great majority of the available tests concentrate on the null of non-stationarity, there are some exceptions, such as the procedure proposed by Kiwiatkoski et al. (1992), but there are not many studies in the empirical literature reporting simultaneously tests for the null of non-stationarity and tests for the null of stationarity. Maddala and Kim (1998) conclude their book on unit roots, cointegration and structural change with a chapter on 'Future directions'; the last section of this chapter, entitled 'What is not needed', contains the following statement:

> (...) what we do not need is more unit root tests (each of which uses the Nelson-Plosser data as a Guinea pig) (...)

We agree with such a view and we prefer to concentrate on multivariate modelling of non-stationary time series and to de-emphasize the debate on deterministic versus stochastic trends within the context of univariate models.

1.1. *Dynamic Models and Spurious Regressions*

Let us reconsider our spurious regression for UK stock prices in the context of a dynamic model. We augment the static regression to consider log-stock prices and log-dividends lagged one period. Results in Table 3 show that the spurious regression problem disappears: lagged US dividends are not significant in explaining UK prices, while UK dividends are.

TABLE 3. Dynamic models for UK stock prices

Dependent variable $LPUK_t$, regression by OLS, 1960:1-1998:1				
	Model with US dividends		Model with UK dividends	
	Coefficient	S.E.	Coefficient	S.E
c	0.218	0.084	0.416	0.081
$LPUK_{t-1}$	0.9423	0.024	0.857	0.036
$LDUS_{t-1}$	0.124	0.09		
$LDUK_{t-1}$			0.159	0.062
Trend	-0.0007	0.001	0.000004	0.0006
R^2	0.99		0.99	
S.E.	0.0096		0.0092	

This is an interesting result which leads us to think that in case the problems related to non-stationarity can be solved, dynamic multivariate time-series models are the right foundation for multivariate time-series modelling.

2. Non Stationary Time-Series, Cointegration and Error Correction Models

Consider the log of stock prices and the log of dividends, these are trending variables, and removing a deterministic trend from them does not deliver stationary time-series. See, for example, in the previous Figure (2) the logarithms of UK aggregate dividends and stock prices. The dynamic dividend growth model is built on the assumption that the log of the dividend price is stationary. The log of the dividends and the log of stock prices are non-stationary seires integrated of order 1 (i.e., their first difference is stationary), however there exists a linear combination of them that becomes stationary. In this case we say that the two series are cointegrated with a cointegrating vector (1, -1). In general, we say that two non-stationary series integrated of order d are cointegrated of order b, if there exists a linear combination of them which is integrated of order $d - b$. Cointegration has interesting and important implications in forecasting and it also explains why in our earlier example, spurious results disappeared when dynamic models were estimated.

To understand the implications of cointegration we have to move from the reality of univariate time series models to the reality of multivariate models, where the joint process of several variables is simultaneously modelled. Let us consider the simplest possible multivariate model, i.e., a bivariate model and let us consider the case of the two specific variables to our interest, lp_t, the log of stock prices and ld_t, the log of dividends. We represent the dynamic process as follows:

$$\begin{aligned} lp_t &= a_0 + a_1 lp_{t-1} + a_2 ld_{t-1} + \epsilon_{1t} \\ ld_t &= b_0 + b_1 ld_{t-1} + \epsilon_{2t}. \end{aligned} \tag{2}$$

Note that system (2) is a multivariate generalization of the univariate autoregressive process, i.e., a bivariate restricted VAR(1) (see the lecture slides), than can be re-written as:

$$\mathbf{Y}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{Y}_{t-1} + \boldsymbol{\epsilon}_t$$

$$\mathbf{Y}_t \equiv \begin{bmatrix} lp_t \\ ld_t \end{bmatrix}, \quad \boldsymbol{\epsilon}_t \equiv \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}, \quad \mathbf{A}_0 \equiv \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}, \quad \mathbf{A}_1 \equiv \begin{bmatrix} a_1 & a_2 \\ 0 & b_1 \end{bmatrix}.$$

Cointegration has interesting implications for VAR representations. Consider the realistic case, when our variables are non-stationary, which is obtained simply by setting $b_1 = 1$:

$$\begin{aligned} lp_t &= a_0 + a_1 lp_{t-1} + a_2 ld_{t-1} + \epsilon_{1t} \\ ld_t &= b_0 + ld_{t-1} + \epsilon_{2t}. \end{aligned}$$

This way, ld_t becomes a random walk with drift; because lp_t is a linear function of a random walk, it becomes itself a random walk. Note first that the short-run elasticity of prices to dividends is different from the long-run elasticity. The short-run elasticity is a_2 while the long-run elasticity is $a_2/(1-a_1)$. The latter is found by setting all variables in the dynamic model (2) to their steady-state value $lp_{t+i} = \overline{lp}$, $ld_{t+i} = \overline{ld}$. To see this point immediately, consider the following re-parameterization of (2):

$$\begin{aligned} \Delta lp_t &\equiv lp_t - lp_{t-1} = a_0 + (a_1 - 1) lp_{t-1} + a_2 ld_{t-1} + \epsilon_{1t} \\ &= a_0 + (a_1 - 1)\left[lp_{t-1} - \left(-\frac{a_2}{a_1 - 1} \right) ld_{t-1} \right] + \epsilon_{1t} \\ \Delta ld_t &\equiv ld_t - ld_{t-1} = b_0 + \epsilon_{2t}. \end{aligned}$$

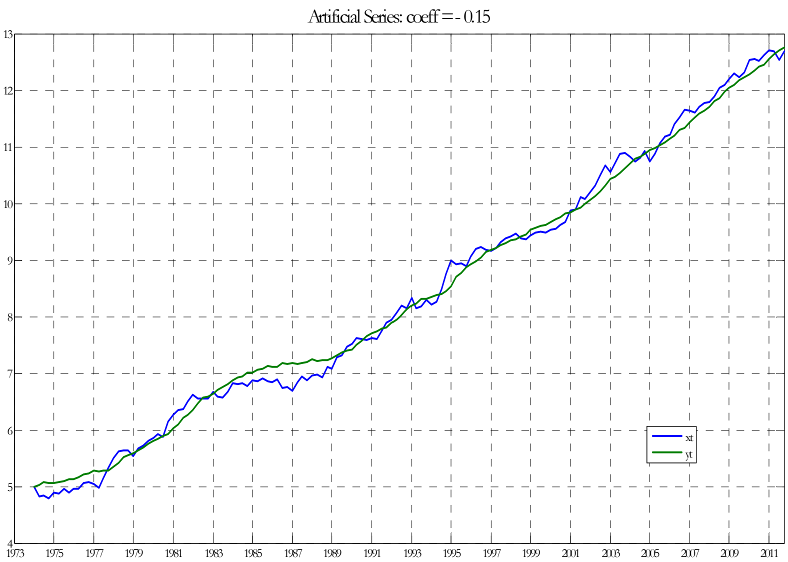
or

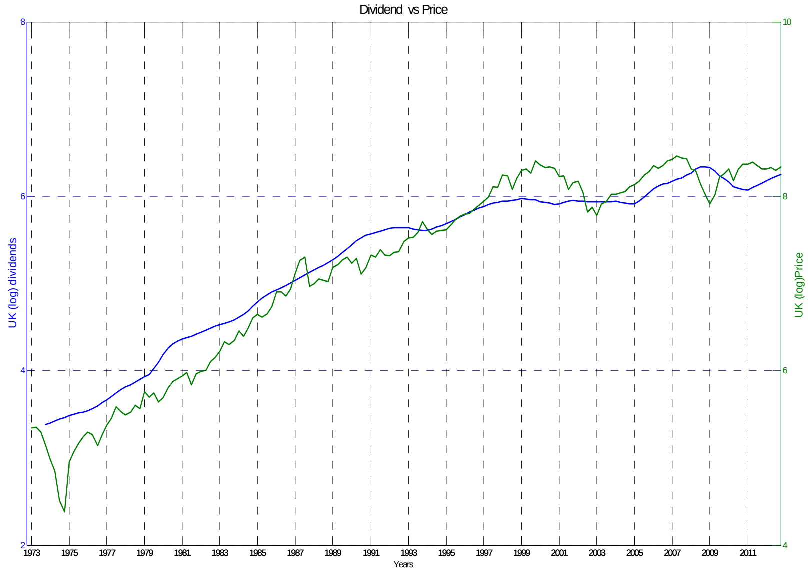
$$\begin{aligned} \Delta lp_t &= a_0 + \alpha \left(lp_{t-1} - \beta_1 ld_{t-1} \right) + \epsilon_{1t} \\ \Delta ld_t &= b_0 + \epsilon_{2t} \\ \alpha &\equiv (a_1 - 1) \qquad \beta_1 \equiv -\frac{a_2}{a_1 - 1}. \end{aligned} \tag{3}$$

The estimated dynamic model includes both first differences and levels. The presence of the level variables generates a long-run solution, derived by setting all first differences either to zero (steady state with no deterministic trend) or a constant (steady state). Note the role of the terms in level: we can interpret $\beta_1 ld_{t-1}$ as the long-run equilibrium level $lp^* \equiv \beta_1 ld_{t-1}$ for the log of prices. When $\alpha < 0$, prices increase at time t whenever $lp_{t-1} < lp^*_{t-1}$, and decreases whenever $lp > lp^*_{t-1}$. The system will tend to self-equilibrate in the presence of any disequilibrium (i.e., a discrepancy between lp and lp^*). Such error correction features guarantee that in the long-run, prices will converge to their equilibrium value. For this reason, the specification (3), with $\alpha < 0$, is termed an error correction model (ECM). Note that, in the case of an ECM representation, the difference between lp and lp^* is a stationary series. *This in fact defines cointegration.* Note that cointegration implies an ECM representation, which allows us to re-write a model in levels, involving non-stationary time-series, as a model involving only stationary variables. Such variables are stationary either because they are the first differences of non-stationary variables or because they are stationary linear combinations of non-stationary variables (cointegrating vectors).

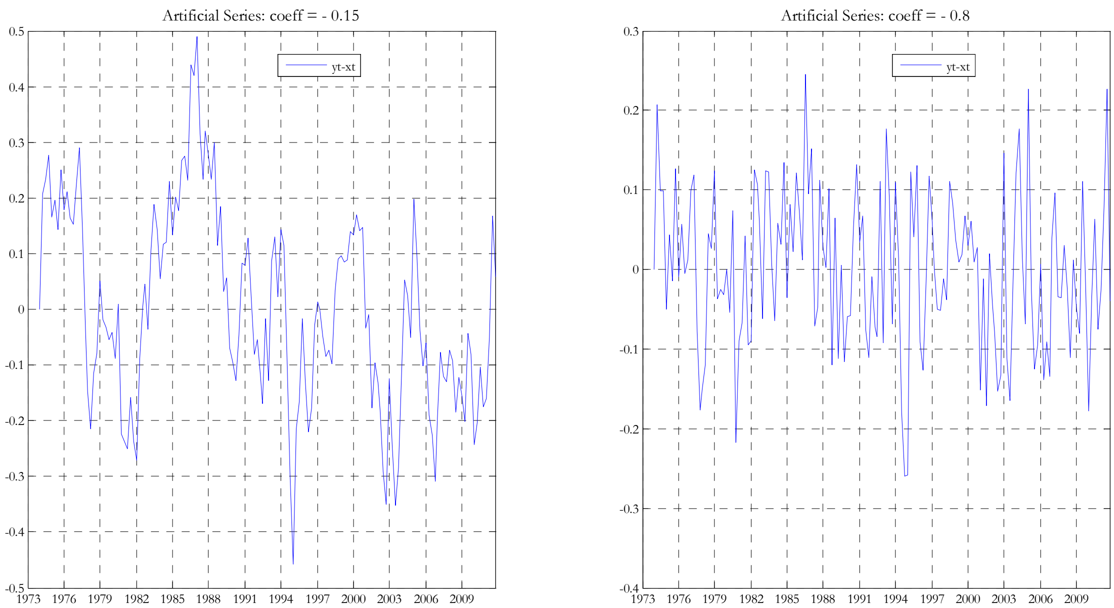
Finally, it is important to emphasize that the prediction of the dividend growth model covered in our past lecture is that $\beta_1 = \frac{a_2}{1-a_1} = 1$. Because Δlp_t represents most of the return variation, cointegration implies that we can use the log of the dividend price ratio as a predictor for stock market returns. This allows us to re-interpret in terms of cointegration between prices and dividends the results of the predictive regressions of stock market returns on the dividend price ratio that we have obtained and commented in earlier chapters.

To show the properties of the model, we first generate samples for the two innovation processes. Then we generate artificial data for (log) prices and dividends by constructing the above model and solving it dynamically. We do so for a sample of 200 observations. The simulated series in levels (lp_t and ld_t) are plotted in the following Figure **??**.





Note that the levels of lp_t and ld_t share a stochastic trend, which disappears from the log dividend yield, the log ratio between the two series. The parameter α in the ECM specification determines the speed of adjustment in the presence of disequilibrium. To illustrate the role of this parameter we report the two series $(lp_t - ld_t)$ generated by taking the same innovations for the sample defined by observations between 1 and 200. The process (**??**) is used to generate the first time series of disequilibrium indicators, $(lp_t - ld_t)$, while the second time-series $(lp_t - ld_t^1)$ is generated by keeping all the parameters unchanged with the exception of α, which is increased from 0.15 to 0.8. The resulting observations for disequilibrium deviations are reported in the following Figure



The disequilibrium deviations in the case of the second simulation are less persistent to support the notion that the second system features a faster speed of adjustment in the presence of deviations from the long-run equilibrium.

2.1. *Static Regressions and Dynamic Models*

As an application of further interest, let us reconsider the static regression in the light of our discussion of dynamic models. Given the following DGP,

$$
\begin{aligned}
y_t &= a_1 y_{t-1} + a_2 x_t + a_3 x_{t-1} + u_{1t} \\
x_t &= b_1 x_{t-1} + u_{2t} \\
\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} &\sim NID\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} \right),
\end{aligned}
\tag{4}
$$

a static model is estimated by OLS:

$$
y_t = \gamma x_t + \varepsilon_t \qquad \widehat{\gamma} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}.
$$

We now assess the results of estimating this static model by taking evaluating the plim $\widehat{\gamma}$ under the true but unknown GDP:

$$
p\lim \widehat{\gamma} = p\lim \left[a_1 \frac{\sum_{t=1}^{T} x_t y_{t-1}/T}{\sum_{t=1}^{T} x_t^2/T} + a_2 + a_3 \frac{\sum_{t=1}^{T} x_t x_{t-1}/T}{\sum_{t=1}^{T} x_t^2/T} + \frac{\sum_{t=1}^{T} x_t u_{1t}/T}{\sum_{t=1}^{T} x_t^2/T} \right].
$$

Under the hypothesis that (4) is stationary ($|b_1| < 1$), we can substitute for x_t in terms of x_{t-1} and u_{2t} and apply Slutsky's and Cramer's theorems to derive the following result:

$$
p\lim \widehat{\gamma} = \frac{a_2 + a_3 b_1}{1 - a_1 b_1} \qquad a_2 \leq p\lim \widehat{\gamma} \leq \frac{a_2 + a_3}{1 - a_1}.
$$

Note that as b_1 approaches zero the elasticity of y_t with respect to x_t delivered by the static regression goes asymptotically to the true short-run elasticity, while as b_1 approaches one, such elasticity converges to the long-run elasticity. Technically speaking, we cannot show what happens when b_1 equals one, because this violates the stationarity conditions which we have used to derive the asymptotic behaviour of the OLS estimator. However, confirming the above intuition, Stock (1987) has shown that the OLS estimator of the parameters determining the long-run relationship of non-stationary cointegrated series is super-consistent. It converges towards the true value at speed of order T, much higher than the speed of order $\sqrt{T}$, with which the OLS estimator converges to its true value in regression involving stationary time series. This result provides the background to a two-step research strategy, according to which the cointegrating relations are estimated first in the static model and then used to estimate a dynamic ECM model, involving only stationary variables. This strategy is less efficient than the simultaneous estimation of short-run and long-run dynamics. The static regression delivers super-consistent estimates of the cointegrating parameters despite being mis-specified, because the omitted variables are the stationary variables determining the short-run dynamics, which, in large samples, should not affect the estimation of cointegrating parameters. However, it must be recognized that research using Monte Carlo simulations has shown that the dimension of the samples required to appeal to the superconsistency property is much higher than that of the samples usually available for time series modelling (see, for example, Banerjee et al., 1986; Banerjee and Hendry, 1992; Banerjee et al., 1993).[1]

3. Spurious Regressions and the Predictability of Returns at Different Frequencies

The evidence of cointegration between log stock prices and log-dividends is not so clear-cut. In fact, the log of the price-dividend ratio is a very persistent time series and the possibility that it contains a unit root cannot be ruled out a priori. As a matter of fact, so far we have used in our empirical analysis the UK dividend price ratio; the evidence from US data speaks less favorably in favour of a mean-reverting (log) dividend-price ratio. A widespread empirical evidence in favor of the dynamic dividend growth model, that supports the stationarity of the log dividend yield, is the one based on multi-period predictive regressions of stock market returns. The performance of the log dividend yield as a predictor of stock market returns improves as the length of the horizon at which returns are defined increases. The following Table illustrates this evidence by reporting the performance of predictive regressions of UK stock returns at one quarter, one year, two years, and three years on the dividend yield.

[1] Moreover, the empirical counterparts of commonly used macroeconomic models is usually represented by dynamic multivariate time series models. Therefore, one pays a price by considering static univariate models as a basis for empirical work.

TABLE 4. Forecasting UK Stock-Market Returns at different horizons

Dependent variable $\sum_{j=1}^{k} r^s_{t+j}$, regression by OLS, 1976:1-2012:4

Horizon	β_0^k	β_1^k	R^2	S.E.	S.E. Dep.Var
1-quarter	0.29 (0.09)	0.08 (0.03)	0.062	0.082	0.084
1-year	1.06 (0.16)	0.301 (0.05)	0.22	0.14	0.163
2-year	1.84 (0.20)	0.52 (0.06)	0.35	0.173	0.220
3-year	2.50 (0.6)	0.70 (0.18)	0.47	0.206	0.285
$\overline{(\text{p} - \text{d})}$				-3.25	

$\sum_{j=1}^{k} r^s_{t+j} = \beta_0^k - \beta_1^k (p_t - d_t) + \varepsilon_{t,t+j}$

k $= 1, 4, 8, 12$

r^s_t are log total real UK stock market returns.

The evidence that long-horizon variables seem to yield significant results where "short-term" approaches have failed, has been questioned. Valkanov(2003) argues that long-horizon regressions will always produce "significant" results, whether or not there is a structural relation between the dependent and independent variables in the regression. This result depends on the fact that a rolling summation of series integrated of order zero behaves asymptotically as a series integrated of order one and, whenever the regressor is persistent, the well-know occurrence of spurious regression between I(1) variables easily emerges. Having established that estimation and testing using long-horizon variables cannot be carried out using the usual regression methods, Valkanov (2003) provides a simple guide on how to conduct estimation and inference using long-horizon regressions. He proposes a rescaled t-statistic, t$/\sqrt{T}$, for testing long-horizon regressions. The asymptotic distribution of this statistic, although non-normal, is easy to simulate and the results are applicable to a general class of long-horizon regressions. In deriving his correction, Valkanov also illustrates that the problem related to spurious regression goes beyond the inadequacy of statistical asymptotic approximation when using overlapping variables. In fact he shows that, even after correcting for serially correlated errors, using the classical Hansen and Hodrick (1980) or Newey-West (1987) standard errors, the small-sample distribution of the estimators and the t-statistics are very different from the asymptotic normal distribution.

To illustrate the Valkanov rescaling procedure consider the following DGP:

$$
\begin{aligned}
r^1_{t+1} &= \alpha + \beta lpd_t + \epsilon_{1t} \qquad (5)\\
(1 + \phi L)\, lpd_t &= \mu + \epsilon_{2t} \\
\phi &= 1 + \frac{c}{T} \qquad \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \sim NID\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11}\, \sigma_{12} \\ \sigma_{12}\, \sigma_{22} \end{bmatrix} \right),
\end{aligned}
$$

where the parameter c measures deviations from the unit root case at a decreasing rate (equal to T) in a neighborhood of 1. The unit-root case corresponds to $c = 0$. The long-horizon variables are defined as

$$r_{t+1}^{k} = \sum_{j=1}^{k} r_{t+1}^{1}.$$

The regression at different horizons is estimated by projecting r_t^k on lpd_t.The simulation of the relevant distribution requires an estimate of the nuisance parameter c. To this end, the long-run restrictions implied by the dynamic dividend growth model can be used.

As shown in chapter 1, the model in (5) implies that one-period total return can be approximated as follows:

$$r_{t+1}^{1} = \rho_0 + \rho lpd_{t+1} + \Delta d_{t+1} - lpd_t. \tag{6}$$

Assuming that the log-dividend follows an autoregressive process,

$$lpd_{t+1} = \phi lpd_t + u_t, \tag{7}$$

and by substituting from (7) into (6) we have that

$$\begin{aligned} r_{t+1}^{1} &= \rho_0 - \beta_1 lpd_t + \varepsilon_{t+1} \\ \varepsilon_{t+1} &= \Delta d_{t+1} + u_t \qquad \beta_1 = (1 - \rho\phi), \end{aligned} \tag{8}$$

where ε_{t+1} is a stationary variable and therefore the $E_t[r_{t+1}^1] = \beta_1 lpd_t$. The k-period horizon return can then be written as follows:

$$r_{t+1}^{k} \approx \tilde{k} - \beta_k x_t + \tilde{\epsilon}_{t+1} \qquad \beta_k = (1 - \rho\phi) \sum_{i=0:}^{k-1} \phi^i.$$

Now, notice that we can write

$$\beta_k = (1 - \rho\phi) \frac{1 - \phi^k}{1 - \phi}. \tag{9}$$

Clearly, if $\rho = 1$, then $\beta_k = 1 - \phi^k$. At this point, recall $\phi = 1 + \frac{c}{T}$ and we can express k in terms of the total length of the available sample as $k = \lfloor \lambda T \rfloor$, from which $T \approx \frac{k}{\lambda}$. Therefore

$$\begin{aligned} \beta_k = 1 - \left(1 + \frac{c}{T}\right)^k &= 1 - \left(1 + \frac{c\lambda}{k}\right)^k \\ \lim_{k \longrightarrow \infty} \left(1 + \frac{c\lambda}{k}\right)^k &= e^{c\lambda} \\ \lim_{k \longrightarrow \infty} \beta_k = \lim_{T \longrightarrow \infty} \beta_{\lfloor \lambda T \rfloor} &= 1 - e^{c\lambda}. \end{aligned} \tag{10}$$

Because we can estimate β_k consistently, we can also find a consistent estimate of c by using the transformation:

$$c^{CONSISTENT} = \frac{1}{\lambda} \log (1 - \beta_k).$$

Given the knowledge of c and λ, the model can be simulated under the null to obtain the critical values of the Valkanov t-statistics. In the case of the models estimated in the earlier Table, the following results are obtained:

TABLE 5. Valkanov-adjusted tests

Dependent variable $\sum_{j=1}^{k} r_{t+j}^{s}$, regression by OLS, 1976:1-2012:4		
Horizon	β_1^k	$t/\sqrt{T}$
1-quarter	0.08 (0.03)	0.25
1-year	0.301 (0.05)	0.51
2-year	0.52 (0.06)	0.72
3-year	0.70 (0.18)	0.93
$\overline{(\mathrm{p}-\mathrm{d})}$		
$\sum_{j=1}^{k} r_{t+j}^{s} = \beta_0^k - \beta_1^k (p_t - d_t) + \varepsilon_{t,t+j}$		
Quantiles of the Valkanov distribution		
0.90	0.43	
0.95	0.54	
0.99	0.71	
r_t^s are log total real UK stock market returns		

The empirical literature on predictability also casts doubts on the validity of the cointegrating relationships between dividends and prices and different models have been proposed based on alternative cointegrating relationships (see, for example the FED model by Lander et al., 1997, or Lettau and Ludgvison's model, 2004). The instability of parameter estimates in econometric models has generated alternative approaches based on stationary representations of the return dynamics (Ferreira and Santa Clara, 2011).

4. **Cointegration with Multiple Cointegrating Vectors**

Until now we have considered cointegration in a bivariate context. Things differ, though, in the multivariate case. In general, among N non-stationary series we may have up to $N-1$ cointegrating vectors; moreover, the single equation dynamic modelling can cause serious troubles when there are multiple cointegrating vectors. To illustrate the problem, let us consider the case of an econometrician who uses cointegration techniques to investigate simultaneously yields on long-term bonds, short term bonds and the stock market. The dynamic dividend growth model delivers one cointegrating equilibrium for the stock market in that $(lp_t - ld_t)$ is stationary; similarly using the linearized definition of bond returns illustrated in earlier chapters, we have that the term spread ($S_t \equiv R_{t,T} - r_t$) is stationary, where $R_{t,T}$ is the long-term bond return with maturity $T > t+1$.

Consider now the case of an econometrician who uses cointegration techniques on bond, bill and stock returns. A dynamic linear model for stock market returns is specified as follows:

$$lp_t = a_0 + a_1 lp_{t-1} + a_2 ld_{t-1} + a_3 R_{t-1,T} + a_4 r_{t-1} + \epsilon_{1t}.$$

This statistical model fits the data well. As it is found that $a_1 < 1$, the investigation leads to the identification of a long-run equilibrium stock market price, which results clearly from the ECM reparameterization of the dynamic model:

$$\begin{aligned} \Delta lp_t &= a_0 + (a_1 - 1)\left[lp_{t-1} - (lp)^*_{t-1}\right] + u_t \\ lp^*_{t-1} &= \frac{a_2}{1-a_1} ld_{t-1} + \frac{a_3}{1-a_1} R_{t-1,T} + \frac{a_4}{1-a_1} r_{t-1}. \end{aligned} \tag{11}$$

This ECM representation might lead to interpret the estimated equation in terms of a dynamic equilibrium for stock prices driven by dividends and the term structure of interest rates. As a matter of fact the variables considered might admit two cointegrating relationship, one capturing the stock market dynamics and the other the bond market dynamics. In this case we would have two equilibrium relationships ($lp_t - ld_t$) and ($R_{t,T} - r_t$). The stock market dynamics could be such that returns react to both disequilibria. This evidence may in principle be consistent with the dynamic dividend growth model if the term spread in the bond market is a leading indicator for future dividend growth. In this case we would have:

$$\Delta lp_t = a_0 + (a_1 - 1)\left[lp_{t-1} - ld_{t-1}\right] + a_3\left(R_{t-1,T} - r_{t-1}\right) + u_t. \tag{12}$$

The statistical specification of (11) and (12) is very similar but their interpretation is very different. This illustrative example illustrates the presence of an identification problem: there are two different structural interpretation of the same reduced form evidence.

The solution of this identification problem requires a framework to allow the researcher to find the number of cointegrating vectors among a set of variables and to identify them. The procedure proposed by Johansen (1988; 1992) within the framework of the vector autoregressive model achieves both results.

So far, we have stressed the importance of the magnitude of the adjustment parameter α as the relevant discriminant to decide on cointegration, but we have not yet provided a statistical framework to test such a hypothesis. We also mentioned the importance of dimensionality of the system in empirical work. In this Section we shall elaborate on these points and illustrate Johansen's approach (1988; 1995) to cointegration in a multivariate framework.

4.1. *Johansen's procedure*

Consider the multivariate generalization of the single-equation dynamic model discussed above, i.e., a vector autoregressive model (VAR) for the vector of, possibly non-stationary,

N-variables $\mathbf{y}_t$:

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + ... + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t. \tag{13}$$

By proceeding in the same way we did for the simple single-equation dynamic model, we can re-parameterize the VAR(p) in levels as a model involving levels and the first differences of variables. Start by subtracting $\mathbf{y}_{t-1}$ from both sides of the VAR to obtain:

$$\Delta\mathbf{y}_t = (\mathbf{A}_1 - \mathbf{I}_N)\,\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + ... + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t. \tag{14}$$

Subtracting and adding $(\mathbf{A}_1 - \mathbf{I})\,\mathbf{y}_{t-2}$ to the right hand side, we have:

$$\Delta\mathbf{y}_t = (\mathbf{A}_1 - \mathbf{I}_N)\,\Delta\mathbf{y}_{t-1} + (\mathbf{A}_1 + \mathbf{A}_2 - \mathbf{I})\,\mathbf{y}_{t-2} + ... + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t. \tag{15}$$

By repeating this procedure until $p-1$, we end up with the following specification:

$$\Delta\mathbf{y}_t = \mathbf{\Pi}_1\Delta\mathbf{y}_{t-1} + \mathbf{\Pi}_1\Delta\mathbf{y}_{t-2} + ... + \mathbf{\Pi}\mathbf{y}_{t-p} + \mathbf{u}_t = \sum_{i=1}^{n-1}\mathbf{\Pi}_i\Delta\mathbf{y}_{t-i} + \mathbf{\Pi}\mathbf{y}_{t-p} + \mathbf{u}_t,$$

where:

$$\mathbf{\Pi}_i = -\left(\mathbf{I}_N - \sum_{j=1}^{i}\mathbf{A}_j\right) \qquad \mathbf{\Pi} = -\left(\mathbf{I}_N - \sum_{i=1}^{n}\mathbf{A}_i\right).$$

Clearly the long-run properties of the system are described by the properties of the matrix $\mathbf{\Pi}$. There are three cases of interest:

1. rank $(\mathbf{\Pi}) = 0$; the system is non-stationary, with no cointegration between the variables considered; this is the only case in which non-stationarity is correctly removed simply by taking the first differences of the variables;

2. rank $(\mathbf{\Pi}) = N$, full; the system is stationary;

3. rank $(\mathbf{\Pi}) = k < N$; the system is non-stationary but there are k cointegrating relationships among the N variables. In this case $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ is an $(N \times k)$ matrix of weights and $\boldsymbol{\beta}$ is an $(N \times k)$ matrix of parameters determining the cointegrating relationships.

Therefore, the rank of $\mathbf{\Pi}$ is crucial in determining the number of cointegrating vectors. Johansen's procedure is therefore based on the fact that the rank of a matrix equals the number of its characteristic roots that differ from zero. Here is the intuition on how the test can be constructed. Having obtained estimates for the parameters in the $\mathbf{\Pi}$ matrix, we associate to them estimates for the N characteristic roots and we order them as follows: $\lambda_1 > \lambda_2 > ... > \lambda_N$. If the variables are not cointegrated, then the rank of $\mathbf{\Pi}$ is zero and all the characteristic roots equal zero. In this case each of the expression $\ln(1-\lambda_i)$ equals zero, too (see below for why these expressions may be useful). If, instead, the rank of $\mathbf{\Pi}$ is one, and $0 < \lambda_1 < 1$, then $\ln(1-\lambda_1)$ is negative and $\ln(1-\lambda_2) = \ln(1-\lambda_3) = ... = \ln(1-\lambda_N) = 0$.

Johansen derives a test on the number of characteristic roots that are different from zero by considering the two statistics:

$$\lambda_{\text{trace}}(k) \equiv -T \sum_{i=k+1}^{N} \ln(1-\hat{\lambda}_i)$$
$$\lambda_{\max}(k, k+1) \equiv -T \ln(1-\hat{\lambda}_{k+1}),$$

where T is the number of observations used to estimate the VAR. The first statistic tests the null of at most k cointegrating vectors against a generic alternative. The test should be performed sequentially starting from the null of at most zero cointegrating vectors up to the case of at most m cointegrating vectors. The second statistic tests the null of at most k cointegrating vectors against the alternative of at most $k+1$ cointegrating vectors. Both statistics are small under their null hypotheses. Critical values are tabulated by Johansen and they depend on the number of non-stationary components under the null and on the specification of the deterministic component of the VAR. Johansen (1994) himself has shown in the past some preference for the trace test, based on the argument that the maximum eigenvalue test does not give rise to a coherent testing strategy, as the initial value k from which the test is initialized may affect the overall outcome within a sequence of tests.

To briefly illustrate the intuition behind the procedure, consider the VAR representation of our simple dynamic model (4), introduced in one of the previous sections, for the two variables x_t and y_t:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{16}$$

The system in (16) can be reparameterized as follows to yield a VECM representation:

$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} a_{11}-1 & a_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \tag{17}$$

from which we have:

$$\mathbf{\Pi} = \begin{bmatrix} a_{11}-1 & a_{12} \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} a_{11}-1 \\ 0 \end{bmatrix}, \quad \boldsymbol{\beta}' = \begin{bmatrix} 1 & -\frac{a_{12}}{1-a_{11}} \end{bmatrix}.$$

Let us now consider the case when we have more than two variables and work our example on the bond and stock market from the previous section. The baseline VAR can be specified as:

$$\begin{bmatrix} lp_t \\ ld_t \\ R_{t,T} \\ r_t \end{bmatrix} = \mathbf{A}_0 + \mathbf{A}_1 \begin{bmatrix} lp_{t-1} \\ ld_{t-1} \\ R_{t-1,T} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix},$$

which could then be reparameterized in VECM form as:

$$\begin{bmatrix} \Delta lp_t \\ \Delta ld_t \\ \Delta R_{t,T} \\ \Delta r_t \end{bmatrix} = \mathbf{\Pi}_0 + \mathbf{\Pi} \begin{bmatrix} lp_{t-1} \\ ld_{t-1} \\ R_{t-1,T} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix}.$$

Because we know that there are two cointegrating vectors, we have:

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' \qquad \text{rank}(\mathbf{\Pi}) = 2 \qquad \boldsymbol{\beta}' = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

A possible specification for $\boldsymbol{\alpha}$ is :

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \\ 0 & \alpha_{32} \\ 0 & 0 \end{bmatrix}.$$

With the above specification for the loadings, stock market prices adjusts both to the presence of disequilibrium conditions in the stock and bond markets, long term bonds react to the spread, while short-term rates and dividends do not respond to disequilibrium conditions:

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \\ 0 & \alpha_{32} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} \alpha_{11} & -\alpha_{11} & \alpha_{12} & -\alpha_{12} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_{32} & -\alpha_{32} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

4.2. *Identification of multiple cointegrating vectors*

Johansen's procedure allows us to identify the number of cointegrating vectors. However, in the case of existence of multiple cointegrating vectors, an interesting identification problem arises: $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are only determined up to the space spanned by them. Thus, for any non-singular matrix Λ conformable by product, we have:

$$\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}\Lambda^{-1}\Lambda\boldsymbol{\beta}'.$$

In other words $\boldsymbol{\beta}$ and $\boldsymbol{\beta}'\Lambda$ are two observationally equivalent bases of the cointegrating space. The obvious implication is that before solving such an identification problem no meaningful economic interpretation of coefficients in cointegrating vectors can be proposed. The solution is imposing a sufficient number of restrictions on parameters such that the matrix satisfying such constraints in the cointegrating space is unique. Such a criterion is derived by Johansen (1992) and discussed in the works of Johansen and Juselius (1990), Giannini (1992) , and Hamilton (1994). Given the matrix of cointegrating vectors $\boldsymbol{\beta}$, we can formulate linear constraints on the different cointegrating vectors using the Υ_i matrices of dimensions $\gamma_i \times N$. Let us consider the columns of $\boldsymbol{\beta}$, i.e. the parameters in each cointegrating vector, ignoring the normalization constraint to one of one variable in each cointegrating vector. Any structure of linear constraints can be represented as

$$\Upsilon_i \boldsymbol{\beta}_i = 0,$$
$$\Upsilon_i \text{ is } (\gamma_i \times N) \quad \boldsymbol{\beta}_i \text{ is } (N \times 1) \quad \text{rank}(\Upsilon_i) = \gamma_i.$$

The same constraints can be expressed in explicit form as

$$\boldsymbol{\beta}_i = \mathbf{S}_i \boldsymbol{\theta}_i,$$

where $\mathbf{S}_i$ is a $N \times (N - \gamma_i)$ matrix, $\boldsymbol{\beta}_i$ is a $N \times 1$ vector, $\boldsymbol{\theta}_i$ is $(N - \gamma_i) \times 1$ vector, $\text{rank}(S_i) = n - \gamma_i$, and $\mathbf{R}_i \mathbf{S}_i = \mathbf{O}$.

A necessary and sufficient condition for identification of parameters in the ith cointegrating vector is:

$$\text{rank } (\mathbf{R}_i \boldsymbol{\beta}) = r - 1. \tag{18}$$

When (21) is satisfied, it is not possible to replicate the ith cointegrating vector by taking linear combinations of the parameters in the other cointegrating vectors. In this case, the matrix obtained by applying to the cointegrating space the restrictions of the ith cointegrating vector will have rank $r - 1$. A necessary condition for identification is immediately derived from the fact that $\mathbf{R}_i \boldsymbol{\beta}$ must have enough rows to satisfy condition (21); therefore, a necessary condition is that each cointegrating vector has at least $r - 1$ restrictions.

A sufficient condition for identification is provided by Johansen by considering the implicit and explicit form of expressing constraints:

Theorem 1 *The ith cointegrating vector is identified by the constraints $\mathbf{S}_1$, $\mathbf{S}_2$, ... , $\mathbf{S}_r$, if for each $k = 1, ..., r - 1$ and for each set of indices $1 < j_1 < ... < j_k < r$ not containing i, we have that rank $[R_i S_{j_1}, ..., R_i S_{j_k}] > k$*

Given identification of the system, we can distinguish the case of just-identification and over-identification. In case of over-identification, the over-identifying restrictions are testable.

To illustrate the procedure, let us reconsider our earlier example. Adopting the following vector representation of the series: $[lp_t \, ld_t \, R_{t,T} \, r_t]'$, and leaving aside normalizations, the matrix $\boldsymbol{\beta}$ can be represented as:

$$\begin{pmatrix} \beta_{11} & 0 \\ -\beta_{11} & 0 \\ 0 & \beta_{32} \\ 0 & -\beta_{42} \end{pmatrix}.$$

Given the following general representation of the matrix $\boldsymbol{\beta}$:

$$\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \\ \beta_{41} & \beta_{42} \end{pmatrix},$$

our constraints imply the following specification for the matrices $\mathbf{R}_i$ and $\mathbf{S}_i$:

$$\mathbf{R}_1 = \begin{pmatrix} 1\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{pmatrix}, \quad \mathbf{S}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix},$$

$$\mathbf{R}_2 = \begin{pmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,1 \end{pmatrix}, \quad \mathbf{S}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}.$$

The necessary conditions for identification are obviously satisfied, while the sufficient conditions for identification require that rank $(\mathbf{R}_1\mathbf{S}_2) \geq 1$, and rank $(\mathbf{R}_2\mathbf{S}_1) \geq 1$. These are also satisfied as

$$\mathbf{R}_1\mathbf{S}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad \mathbf{R}_2\mathbf{S}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$$

4.3. *Hypothesis testing with multiple cointegrating vectors*

Johansen's procedure allows for testing the validity of restricted forms of cointegrating vectors. More precisely, the validity of restrictions (over-identifying restrictions) in addition to those necessary to identify the long-run equilibria can be tested. The intuition behind the construction of all tests is that when there are r cointegrating vectors, only exactly r linear combination of variables are stationary; therefore, the test statistics involve comparing the number of cointegrating vectors under the null and the alternative hypotheses. Following this intuition, we understand why only the over-identifying restrictions can be tested. Just-identified models feature the same long-run matrix $\mathbf{\Pi}$, and therefore, the same eigenvalues of $\mathbf{\Pi}$. Consider the case of testing restrictions on a set of r identified cointegrating vectors stacked in the matrix $\boldsymbol{\beta}$. The test statistic involves comparing the number of cointegrating vectors under the null and the alternative hypothesis. Let $\hat{\lambda}_1$, $\hat{\lambda}_2$, ..., $\hat{\lambda}_r$ be the ordered eigenvalues of the $\mathbf{\Pi}$ matrix in the unrestricted model, and $\hat{\lambda}_1^*$, $\hat{\lambda}_2^*$, ..., $\hat{\lambda}_r^*$ the ordered eigenvalues of the $\mathbf{\Pi}$ matrix in the restricted model. Restrictions on $\boldsymbol{\beta}$ are testable by forming the following test statistic:

$$T\sum_{i=1}^{r}\left[\ln\left(1-\hat{\lambda}_i^*\right)-\ln\left(1-\hat{\lambda}_i\right)\right]. \qquad (19)$$

Johansen (1992) shows that the statistic in (19) has a χ^2 distribution with a number of degrees of freedom equal to the number of over-identifying restrictions. Note that small values of $\hat{\lambda}_i^*$ when compared to $\hat{\lambda}_i$ imply a reduction of the rank of $\mathbf{\Pi}$ when the restrictions are imposed and hence the rejection of the null hypothesis. This testing procedure can be extended to tests of restrictions on the matrix of weights $\boldsymbol{\alpha}$ or on the deterministic components (constant and trends) of the cointegrating vectors.

5. Using VAR Models

A Cointegrated VAR, after the identification of the number and shape of cointegrating vector(s), provides a statistical model of the joint distribution of the variables of interests:

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{y}_{t-1} + \mathbf{u}_t \qquad \mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{20}$$

where $\mathbf{y}_t$ is a vector of length N containing the variables under investigation. The reduced form specification (20) can be adopted directly for forecasting purposes or to describe the dynamic response of the system to innovations to observables, such as the VAR residuals. Some further *identification* choice must be made if the model is to be used for evaluating the response of economic and financial variables to innovations to unobservables, i.e., the "structural" shocks to some of the variables included in the VAR. Impulse response analysis examines the effect of a typical shock, usually one-standard deviation perturbation, on the time path of the variables in the model.

Computing impulse responses to unobservables requires the imposition of some identification assumptions and the orthogonality of structural shocks is a necessary condition to consider the effect of each identified shocks in isolation. The study of the response to the system to an innovation in observables does not require any identification assumptions, even though the contemporaneous linkages between shocks must be modelled.[2] In finance, the use of VARs is more related to forecasting first and second moments of the distribution of asset returns at different horizons. Macro-finance models focus on the different role of permanent versus transitory shocks to understand the comovement between financial and macroeoconomic variables.

5.1. *Identification of VARs*

Given the estimate of (20), the problem of extracting unobservable structural shocks $\boldsymbol{\upsilon}_t$ from the observed VAR innovations $\mathbf{u}_t$ is usually addressed by positing the following relations

$$\mathbf{A}\mathbf{u}_t = \mathbf{B}\boldsymbol{\upsilon}_t \qquad \boldsymbol{\upsilon}_t \sim N(\mathbf{0}, \mathbf{I}_N),$$

[2] In macroeconomics, the importance of computing impulse responses to structural shocks is related to the fact that the solution of a Dynamic Stochastic General Equilibrium (DSGE) model can be well approximated by a VAR, and VARs have become the natural tool for model evaluation. In this context, VAR models are not estimated to yield advice on the best policy but rather to provide empirical evidence on the response of macroeconomic variables to policy impulses in order to discriminate between alternative theoretical models of the economy. It then becomes crucial to identify policy actions using restrictions independent from the theoretical models of the transmission mechanism under empirical investigation, taking into account the potential endogeneity of policy instruments.

from which we can derive the relation between the variance-covariance matrices of $\mathbf{u}_t$ (observed) and $\boldsymbol{\nu}_t$ (unobserved) as follows:

$$E\left(\mathbf{u}_t\mathbf{u}_t'\right) = \mathbf{A}^{-1}\mathbf{B}E\left(\boldsymbol{\upsilon}_t\boldsymbol{\upsilon}_t'\right)\mathbf{B}'\mathbf{A}^{-1}.$$

Substituting population moments with sample moments we have:

$$\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\mathbf{I}_N\hat{\mathbf{B}}'\hat{\mathbf{A}}^{-1}, \tag{21}$$

where $\hat{\boldsymbol{\Sigma}}$ contains $N(N+1)/2$ different elements, which is the maximum number of identifiable parameters in matrices **A** and **B.** Therefore a necessary condition for identification is that the maximum number of parameters contained in the two matrices equals $N(N+1)/2$, because such a restriction makes the number of equations equal to the number of unknowns in system (21). As usual, for such a condition also to be sufficient for identification no equation in (21) should be a linear combination of the other equations in the system (see Amisano and Giannini 1996, Hamilton 1994). As in the case of standard VAR models, we have the three possible cases of under-identification, exact-identification, and over-identification. The validity of over-identifying restrictions can be tested via a statistic distributed as a χ^2 with a number of degrees of freedom equal to the number of the over-identifying restrictions that have been imposed. Once identification has been achieved, the estimation problem is solved by applying a generalized method of moments estimation. In practice, identification requires the imposition of some restrictions on the parameters of **A** and **B.** This step has been historically implemented in a number of different ways.

5.2. *Identification of VAR models*

After the identification of structural shocks of interest, the properties of VAR models are described using impulse response analysis, variance decompositions, and historical decompositions. Consider a structural VAR model for a generic vector $\mathbf{y}_t$, containing N variables. Given an identified and estimated structural VAR(p),

$$\mathbf{y}_t = \sum_{i=1}^{p}\mathbf{C}_i\mathbf{y}_{t-i} + \mathbf{u}_t \qquad \mathbf{A}\mathbf{u}_t = \mathbf{B}\boldsymbol{\upsilon}_t,$$

we can re-write it as

$$\mathbf{A}\mathbf{y}_t = \sum_{i=1}^{p}\mathbf{A}_i\mathbf{y}_{t-i} + \mathbf{B}\boldsymbol{\upsilon}_t \qquad \mathbf{A}^{-1}\mathbf{A}_i = \mathbf{C}_i,$$

which we can express in a compact fashion as:

$$\left[\mathbf{A} - \mathbf{A}\left(L\right)\right]\mathbf{y}_t = \mathbf{B}\mathbf{v}_t$$
$$\mathbf{A}\left(L\right) = \sum_{i=1}^{p}\mathbf{A}_iL^i.$$

By inverting $[\mathbf{A}_0 - \mathbf{A}(L)]$ (under the assumption of invertibility of this polynomial) we obtain the moving average representation for our VAR process:

$$\mathbf{y}_t = \mathbf{C}(L)\,\mathbf{v}_t = \mathbf{C}_0\mathbf{v}_t + \mathbf{C}_1\mathbf{v}_{t-1} + ... + \mathbf{C}_s\mathbf{v}_{t-s} \tag{22}$$

where $\mathbf{C}(L) = [\mathbf{A}_0 - \mathbf{A}(L)]^{-1}$ and $\mathbf{C}_0 = \mathbf{A}_0^{-1}\mathbf{B}$. To illustrate the concept of an *impulse response function,* we interpret the generic matrix $\mathbf{C}_s$ within the moving average representation as follows:

$$\mathbf{C}_s = \frac{\partial \mathbf{y}_{t+s}}{\partial \mathbf{v}_t}.$$

The generic element $[i,j]$ of the matrix $\mathbf{C}_s$ represents the impact of a shock hitting the jth variable of the system at time t on the ith variable of the system at time $t+s$. As s varies we have a function describing the response of variable i to an impulse in variable j. For this function of partial derivatives to be meaningful we must allow that a shock to variable j occurs while all other shocks are kept to zero. Of course this is natural for structural shocks, as they are identified by imposing they are orthogonal to each other. Note, however that the concept of an impulse response function is not applicable to reduced form VAR innovations, which, in general, are correlated to each other.

A *historical decomposition* is obtained by using the structural MA representation to separate series in the components (orthogonal to each other) attributable to the different structural shocks.

Finally, a *forecast error variance decomposition* (FEVD) is obtained from (22) by deriving the error in forecasting $\mathbf{y}_s$ period in the future as

$$\mathbf{y}_{t+s} - E_t[\mathbf{y}_{t+s}] = \mathbf{C}_0\mathbf{v}_t + \mathbf{C}_1\mathbf{v}_{t-1} + ... + \mathbf{C}_s\mathbf{v}_{t-s},$$

from which we can construct the variance of such forecast errors as:

$$Var\left(\mathbf{y}_{t+s} - E_t[\mathbf{y}_{t+s}]\right) = \mathbf{C}_0\mathbf{I}_N\mathbf{C}_0' + \mathbf{C}_1\mathbf{I}_N\mathbf{C}_1' + ... + \mathbf{C}_s\mathbf{I}_N\mathbf{C}_s'$$

from which we can compute the share of the total variance attributable to the variance of each structural shock. Note again that such decomposition makes sense only if the shocks are orthogonal to each other. Only in this case, we can write the variance of the total forecast error as a sum of the variances of the individual shocks (as the covariance terms are zero following the orthogonality property of structural shocks). In practice, identification requires the imposition of some restrictions on the parameters of the matrices $\mathbf{A}$ and $\mathbf{B}$. This step has been historically implemented in a number of different ways.

5.3. *Choleski decomposition*

In the famous article which introduced the VAR methodology to the profession, Sims (1980*a*) proposed the following identification strategy, based on the Choleski decomposition of the of matrices:

$$\mathbf{A} = \begin{pmatrix} 1 & 0\,0 & & 0 \\ a_{21} & 1\,0 & & 0 \\ . & .\,1 & & . \\ a_{n1} & . & a_{nn-1} & 1 \end{pmatrix}, \qquad \mathbf{B} \begin{pmatrix} b_{11} & 0 & 0 & 0 \\ 0 & b_{22} & 0 & 0 \\ . & . & b_{ii} & . . \\ 0 & 0 & 0 & b_{nn} \end{pmatrix}. \tag{23}$$

This is obviously an exact identification scheme, where the identification of structural shocks depends on the ordering of variables. It corresponds to a recursive economic structure, with the most endogenous variable ordered last.

Consider for the sake of illustration a bivariate VAR:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}.$$

Its MA representation is

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t-1} \\ v_{2t-1} \end{pmatrix}$$
$$+ ... + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^s \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t-s} \\ v_{2t-s} \end{pmatrix},$$

from which impulse response functions, historical decomposition and forecasting error variance decomposition are immediately obtained. An obvious generalization of Choleski is to consider contemporaneous restrictions that do not necessarily lead to a triangular structure of A.

5.4. *CVAR and the identification of shocks*

Consider, for simplicity, the case of a bivariate model $\mathbf{y}_t = [y_t,\ x_t]'$, in which the variables are non-stationary ($I(1)$) but cointegrated with a cointegrating vector $(1, -1)$, so the rank of the $\mathbf{\Pi}$ matrix is 1 and we use the following representation of the stationary reduced form:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{24}$$

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}. \tag{25}$$

Model (24) can be re-written as follows :

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} (1-L) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & 0 \end{pmatrix} \begin{pmatrix} (y_{t-1} - x_{t-1}) \\ \Delta x_{t-1} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}. \tag{26}$$

The two representations are completely identical (they feature the same residuals). The cointegrating properties of the system suggest the presence of two types of shocks: a permanent one (related to the single common trend shared by the two variables) and a transitory one (related to the cointegrating relation). It seems therefore natural to identify one shock as permanent and the other as transitory. Given that we have a stationary system, the identification of shocks is obtained by deriving long-run responses of the variables of interest to relevant shocks. From (26) we have:

$$\left(\begin{pmatrix} 1\,1 \\ 0\,1 \end{pmatrix}\begin{pmatrix} (1-L)\,0 \\ 0 \quad\quad 1 \end{pmatrix} - \begin{pmatrix} \alpha_{11}L\,0 \\ \alpha_{21}L\,0 \end{pmatrix}\right)\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} b_{11}\,b_{12} \\ b_{21}\,b_{22} \end{pmatrix}\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix},$$

from which long-run responses are obtained by setting $L = 1$ and by inverting the matrix pre-multiplying variables in the stationary representation of VAR

$$\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} -\alpha_{11}\,1 \\ -\alpha_{21}\,1 \end{pmatrix}^{-1}\begin{pmatrix} b_{11}\,b_{12} \\ b_{21}\,b_{22} \end{pmatrix}\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{-b_{11}+b_{21}}{\alpha_{11}-\alpha_{21}} & -\frac{b_{12}-b_{22}}{\alpha_{11}-\alpha_{21}} \\ \frac{-\alpha_{21}b_{11}+\alpha_{11}b_{21}}{\alpha_{11}-\alpha_{21}} & \frac{-\alpha_{21}b_{12}+\alpha_{11}b_{22}}{\alpha_{11}-\alpha_{21}} \end{pmatrix}\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}. \qquad (27)$$

Thus v_{2t} can be identified as the transitory shock by imposing the following restriction

$$-\alpha_{21}b_{12} + \alpha_{11}b_{22} = 0,$$

which, given knowledge of the $\boldsymbol{\alpha}$ parameters from the cointegration analysis, provides the just-identifying restriction for the parameters in $\mathbf{B}$. Note that, there is one case in which this identification is equivalent to the Choleski ordering, the case in which $\alpha_{11} = 0$. Note that this is the case in which Δy_t is weakly exogenous for the estimation of b_{21}.[3]

5.5. *Sign Restrictions*

Given the VAR specification,

$$\mathbf{y}_t = \sum_{i=1}^{p}\mathbf{A}_i\mathbf{y}_{t-i} + \mathbf{B}\mathbf{u}_t \qquad \boldsymbol{\Sigma} = \mathbf{B}E[\mathbf{u}_t\mathbf{u}_t']\mathbf{B}' = \mathbf{B}\mathbf{B}'.$$

Consider the Choleski decomposition of $\boldsymbol{\Sigma}$ and $\mathbf{C}$. The impulse response function, given the Choleski decomposition can be written as:

$$\mathbf{y}_t = [\mathbf{I} - \mathbf{A}(L)]^{-1}\mathbf{C}\mathbf{u}_t.$$

All the possible rotations of the Choleski decomposition are then obtained as follows:

$$[\mathbf{I} - \mathbf{A}(L)]^{-1}\mathbf{C}\mathbf{Q}\mathbf{Q}'\mathbf{u}_t$$

where $\mathbf{Q}\mathbf{Q}' = \mathbf{I}_N$. The impulse response for $\mathbf{Q}'\mathbf{u}_t$, is then $[\mathbf{I} - \mathbf{A}(L)]^{-1}\mathbf{C}\mathbf{Q}$**.** The imposition of the sign restrictions then consider $\mathbf{Q}$ to generate all possible identifications and then select only those that satisfy some sign restriction.

[3] An application of this identifying scheme is provided in Cochrane (1999) that uses it to identify permanent and transitory components in stock prices.

5.6. *Generalized impulse response functions*

If the identification of structural shocks is not an issue of primary interest then Generalized Impulse Response Functions can also be used to describe the response of a VAR system to change in observable variables, i.e., to VAR innovations. To provide an example, consider again our bivariate CVAR model:

$$\begin{bmatrix} (y_t - x_t) \\ \Delta x_t \end{bmatrix} = \mathbf{A} \begin{bmatrix} (y_{t-1} - x_{t-1}) \\ \Delta x_{t-1} \end{bmatrix} + \mathbf{u}_t$$
$$\mathbf{u}_t \sim N\left(\mathbf{0}, \begin{bmatrix} \sigma_{11}^2 \, \sigma_{12} \\ \sigma_{12} \, \sigma_{22}^2 \end{bmatrix}\right).$$

From the properties of the normal distribution we have that:

$$E[u_{2t}|u_{1t}] = \frac{\sigma_{12}}{\sigma_{11}^2} u_{1t}$$

so the impulse responses can be simply derived as follows:

$$\frac{\partial \begin{bmatrix} (y_{t+i} - x_{t+i}) \\ \Delta x_t + i \end{bmatrix}}{\partial u_{1t}} = \mathbf{A}^i \mathbf{S} = \begin{bmatrix} 1 \\ \left(\sigma_{11}^2\right)^{-1} \sigma_{12} \end{bmatrix}.$$

The generalized impulse response functions (GIRFs) seem to be more appropriate when the primary focus of the analysis is the description of the transmission mechanism rather than the structural interpretation of shocks. The effect of the shock we are studying with GIRFs can be interpreted as the effect on the variables in the model of an intercept adjustment to the particular equation shocked.

6. **Cointegration and Present Value Models**

Consider a vector $\mathbf{y}_t$ containing two variables x_t and z_t cointegrated with an equilibrium error given by $S_t = x_t - \beta z_t$. Johansen's ECM representation for such system is:

$$\begin{aligned} \begin{bmatrix} \Delta x_t \\ \Delta z_t \end{bmatrix} &= \mathbf{\Pi}_1 \begin{bmatrix} \Delta x_{t-1} \\ \Delta z_{t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} \begin{bmatrix} 1 & -\beta \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \\ &= \mathbf{\Pi}_1 \begin{bmatrix} \Delta x_{t-1} \\ \Delta z_{t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} S_{t-1} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}. \end{aligned} \tag{28}$$

Define a matrix $\mathbf{M}$ such that

$$\mathbf{M} \begin{bmatrix} \Delta x_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} \Delta x_t \\ \Delta S_t \end{bmatrix}$$

so it must be:

$$\mathbf{M} \equiv \begin{bmatrix} 1 & 0 \\ 1 & -\beta \end{bmatrix}.$$

Therefore we have:

$$\mathbf{M}\begin{bmatrix}\Delta x_t\\ \Delta z_t\end{bmatrix} = \mathbf{M\Pi}_1\begin{bmatrix}\Delta x_{t-1}\\ \Delta z_{t-1}\end{bmatrix} + \mathbf{M}\begin{bmatrix}\alpha_{11}\\ \alpha_{21}\end{bmatrix} S_{t-1} + \mathbf{M}\begin{bmatrix}v_{1t}\\ v_{2t}\end{bmatrix}$$
$$\begin{bmatrix}\Delta x_t\\ \Delta z_t\end{bmatrix} = \mathbf{M\Pi}_1\mathbf{M}^{-1}\begin{bmatrix}\Delta x_{t-1}\\ \Delta S_{t-1}\end{bmatrix} + \mathbf{M}\begin{bmatrix}\alpha_{11}\\ \alpha_{21}\end{bmatrix} S_{t-1} + \mathbf{M}\begin{bmatrix}v_{1t}\\ v_{2t}\end{bmatrix}.$$

The system can now be re-arranged so that it describes levels rather than differences of S_t. The result is a second order VAR as follows:

$$\begin{bmatrix}\Delta x_t\\ S_t\end{bmatrix} = \mathbf{G}_1\begin{bmatrix}\Delta x_{t-1}\\ S_{t-1}\end{bmatrix} + \mathbf{G}_2\begin{bmatrix}\Delta x_{t-2}\\ S_{t-2}\end{bmatrix} + \mathbf{M}\begin{bmatrix}v_{1t}\\ v_{2t}\end{bmatrix}.$$

Campbell and Shiller test the expectations hypothesis (EH) of interest rates by considering the case of the risk free rate (r_t) and a very long term bond ($R_{t,T}$) with maturity $T >> t$. In such case under the null of the EH we have:[4]

$$R_{t,T} = R^*_{t,T} \approx (1-\gamma)\sum_{j=0}^{T-t-1}\gamma^j E_t[r_{t+j}] \tag{29}$$

which can be re-written in terms of spread between long and short-term rates, $S_{t,T} = R_{t,T} - r_t$:

$$S_{t,T} = S^*_{t,T} = \sum_{j=1}^{T-t-1}\gamma^j E_t[\Delta r_{t+j}]. \tag{30}$$

(30) shows that a necessary condition for the EH to hold puts constraints on the long-run dynamics of the spread. In fact, the spread should be stationary being a weighted sum of stationary variables. Obviously, stationarity of the spread implies that, if yields are non-stationary, they should be cointegrated with a cointegrating vector $(1, -1)'$. However, the necessary and sufficient conditions for the validity of the EH impose restrictions both on the long-run and the short run dynamics.

Assuming that $R_{t,T}$ and r_t are cointegrated with a cointegrating vector $(1, -1)'$, Campbell and Shiller construct a bivariate stationary VAR in first differences for the short-term rate and the spread:[5]

$$\begin{aligned}\Delta r_t &= a(L)\Delta r_{t-1} + b(L)S_{t-1} + u_{1t}\\ S_t &= c(L)\Delta r_{t-1} + d(L)S_{t-1} + u_{2t}\end{aligned} \tag{31}$$

Stacking the VAR in compact form as,

[4] In fact, CS use de-meaned-variables, that is equivalent to test a weak form of the EH, in the sense that de-meaning eliminates a constant risk premium.

[5] In fact, the evidence for the restricted cointegrating vector which constitutes a necessary condition for the EH to hold is not found to be particularly strong in the original work by Campbell and Shiller.

$$
\underbrace{\begin{bmatrix} \Delta r_t \\ . \\ . \\ \Delta r_{t-p+1} \\ S_t \\ . \\ . \\ S_{t-p+1} \end{bmatrix}}_{\mathbf{z}_t} = \underbrace{\begin{bmatrix} a_1 \,.\,.\, a_p \; b_1 \,.\,.\, b_p \\ 1 \,.\,.\, 0 \;\; 0 \,.\,.\, 0 \\ 0 \,.\,.\, 0 \;\; 0 \,.\,.\, 0 \\ 0 \,.\, 1 \; 0 \;\; 0 \,.\,.\, 0 \\ c_1 \,.\,.\, c_p \; d_1 \,.\,.\, d_p \\ 0 \,.\,.\, 0 \;\; 1 \,.\,.\, 0 \\ 0 \,.\,.\, 0 \;\; 0 \,.\,.\, 0 \\ 0 \,.\,.\, 0 \;\; 0 \,.\, 1 \; 0 \end{bmatrix}}_{\mathbf{A}\mathbf{z}_t} \underbrace{\begin{bmatrix} \Delta r_{t-1} \\ . \\ . \\ \Delta r_{t-p} \\ S_{t-1} \\ . \\ . \\ S_{t-p} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \underbrace{\begin{bmatrix} u_{1t} \\ . \\ . \\ 0 \\ u_{2t} \\ . \\ . \\ 0 \end{bmatrix}}_{\mathbf{v}_t}.
$$

$$
= \mathbf{A}\mathbf{z}_{t-1} + \mathbf{v}_t \tag{32}
$$

The null hypothesis of the EH imposes a set of restrictions which can be written as:

$$
\mathbf{g}'\mathbf{z}_t = \sum_{j=1}^{T-1} \gamma^j \mathbf{h}'(\mathbf{A}^j)'\mathbf{z}_t \tag{33}
$$

where $\mathbf{g}'$ and $\mathbf{h}'$ are appropriate selector vectors for S and Δr, respectively (i.e., row vectors with $2p$ elements, all of which are zero except for the $p+1$th element of $\mathbf{g}'$ and the first element of $\mathbf{h}'$, which are equal to one). Since the above expression has to hold for all $\mathbf{z}_t$s, and, for large T, the sum in (33) converges under the null of the validity of the EH, it must be the case that:

$$
\mathbf{g}' = \mathbf{h}'\gamma\mathbf{A}(\mathbf{I} - \gamma\mathbf{A})^{-1}, \tag{34}
$$

which implies

$$
\mathbf{g}'(I - \gamma\mathbf{A}) = \mathbf{h}'\gamma\mathbf{A} \tag{35}
$$

from which we derive the following constraints on the individual coefficients of the VAR in (2):

$$
\{c_i = -a_i, \forall i\}, \quad \{d_1 = -b_1 + 1/\gamma\}, \quad \{d_i = -b_i, \forall i \neq 1\}. \tag{36}
$$

The above restrictions are testable using a Wald test. By doing so, using US data spanning the period between the 1950s and the 1980s, Campbell and Shiller (1987) rejected the null of the EH. However, when they construct a theoretical spread $S^*_{t,T}$, by imposing the (albeit rejected) EH restrictions on their estimated VAR, they find that, despite the statistical rejection of the EH, $S^{*}_{t,T}$ and $S_{t,T}$ are strongly correlated, as shown in the following figure.

1080 JOURNAL OF POLITICAL ECONOMY

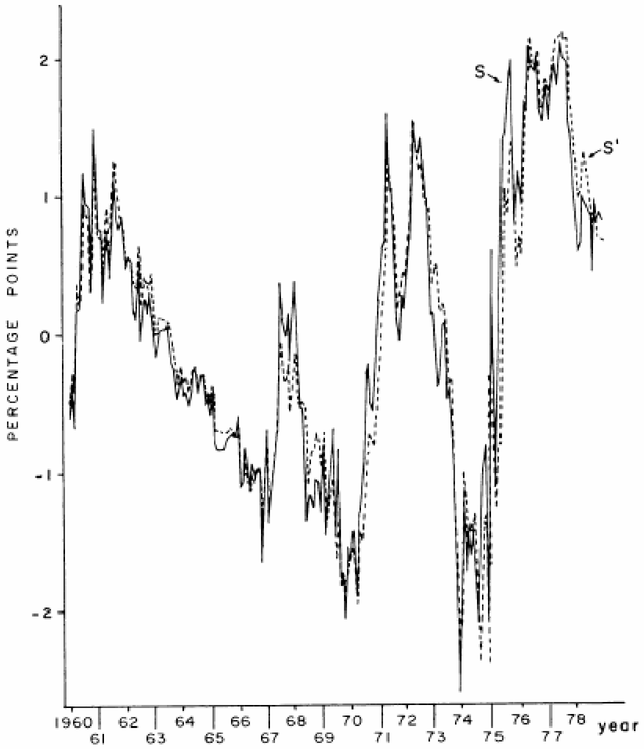


FIG. 1.—Term structure: deviations from means of long-short spread S_t and theoretical spread S'_t.

Things look very different in applications to stock market data, when the dynamic dividend growth model with constant rates of return is considered. In this case we have

$$(p_t - d_t)^* = \sum_{j=1}^{m} \rho^{j-1} \left(\Delta d_{t+j} \right),$$

and the variable $(p_t - d_t)^*$ can be obtained by imposing the appropriate cross-equation restrictions on a bivariate VAR for the dividend-yield and the dividend growth rate. The relation between the actual and the "theory-consistent" price-dividend ratio looks much different than what had been obtained for the bond market, see the following Figure. This result is consistent from the evidence of predictive regressions relating the dividend price to future returns rather than to future dividend growth.

COINTEGRATION 1083

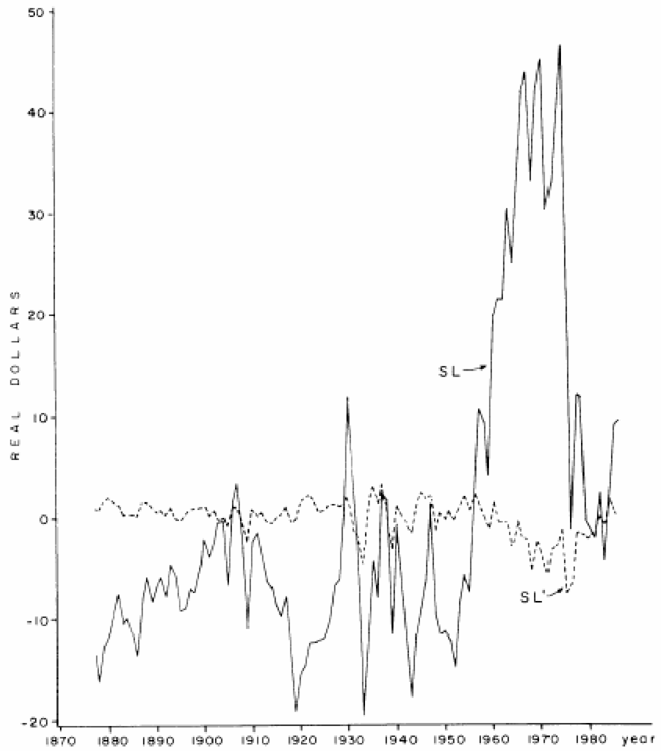


FIG. 2.—Stock market: deviations from means of actual spread (SL_t = Price$_t$ − θ · Dividend$_{t-1}$) and theoretical spread SL'_t, θ = 12.195.

6.1. *Cointegration and multivariate trend-shock decompositions*

Having discussed the VECM representation for a vector of N non-stationary variables admitting k cointegrating relationships, let us compare it with the multivariate extension of the Beveridge-Nelson decomposition. Consider the simple case of an $I(1)$ vector $\mathbf{y}_t$ featuring first-order dynamics and no deterministic component:

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{t-1} + \mathbf{u}_t, \tag{37}$$

where $\boldsymbol{\alpha}$ is the $N \times k$ matrix of loadings and $\boldsymbol{\beta}$ is the $N \times k$ matrix of parameters in the cointegrating relationships. Because $\mathbf{y}_t$ is $I(1)$, we can apply the Wold decomposition theorem to $\Delta \mathbf{y}_t$ to obtain the following representation:

$$\Delta \mathbf{y}_t = \mathbf{C}(L)\, \mathbf{u}_t,$$

from which, by applying the algebra illustrated in our discussion of the univariate Beveridge-Nelson decomposition, we can derive the following stochastic-trend representation:

$$\mathbf{y}_t = \mathbf{C}^*(L)\, \mathbf{u}_t + \mathbf{C}(1)\, \mathbf{z}_t.$$

Here $\mathbf{z}_t$ is an appropriate process for which $\Delta \mathbf{z}_t = \mathbf{u}_t$. The existence of cointegration imposes restrictions on the $\mathbf{C}$ matrices. The stochastic trends must cancel out when the k stationary linear combinations of the variables in $\mathbf{y}_t$ are considered. In other words we must have:

$$\boldsymbol{\beta}' \mathbf{C}(1) = \mathbf{0}.$$

By investigating further the relation between the VECM and stochastic trend representations, we can give a more precise parameterization of the matrix $\mathbf{C}(1)$. First, note that equation (**??**) is equivalent to:

$$\mathbf{y}_t = (\mathbf{I}_N + \boldsymbol{\alpha}\boldsymbol{\beta}')\,\mathbf{y}_{t-1} + \mathbf{u}_t. \tag{38}$$

Pre-multiplying this system by $\boldsymbol{\beta}'$ yields:

$$\begin{aligned}\boldsymbol{\beta}'\mathbf{y}_t &= \boldsymbol{\beta}'(\mathbf{I}_N + \boldsymbol{\alpha}\boldsymbol{\beta}')\,\mathbf{y}_{t-1} + \boldsymbol{\beta}'\mathbf{u}_t \\ &= (\mathbf{I}_k + \boldsymbol{\alpha}\boldsymbol{\beta}')\,\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\beta}'\mathbf{u}_t.\end{aligned}$$

Solving this model recursively, we obtain the MA(∞) representation for the k cointegrating relationships:

$$\boldsymbol{\beta}'\mathbf{y}_t = \sum_{i=0}^{\infty} (\mathbf{I}_k + \boldsymbol{\alpha}\boldsymbol{\beta}')^{i}\,\boldsymbol{\beta}'\mathbf{u}_{t-i}. \tag{39}$$

By substituting (39) in (**??**), we obtain the MA representation for $\Delta\mathbf{y}_t$,

$$\Delta\mathbf{y}_t = \sum_{i=1}^{\infty} \boldsymbol{\alpha}(\mathbf{I}_k + \boldsymbol{\alpha}\boldsymbol{\beta}')^{i-1}\,\boldsymbol{\beta}'\mathbf{u}_{t-i} + \mathbf{u}_t,$$

from which we have

$$\mathbf{C}(1) = \mathbf{I}_N - \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}'. \tag{40}$$

Now note the beautiful relationship (see Johansen 1995, p. 40),

$$\mathbf{I}_N = \boldsymbol{\beta}_\perp(\boldsymbol{\alpha}'_\perp\boldsymbol{\beta}_\perp)^{-1}\boldsymbol{\alpha}'_\perp + \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}', \tag{41}$$

where $\boldsymbol{\beta}_\perp, \boldsymbol{\alpha}_\perp$ are $N \times (N-k)$ matrices of rank $N-k$ such that $\boldsymbol{\alpha}'_\perp\boldsymbol{\alpha} = 0$, $\boldsymbol{\beta}'_\perp\boldsymbol{\beta} = 0$. Using (41) in (40), we have

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

and

$$\mathbf{y}_t = \mathbf{C}^*(L)\,\mathbf{u}_t + \boldsymbol{\beta}_\perp(\boldsymbol{\alpha}'_\perp\boldsymbol{\beta}_\perp)^{-1}(\boldsymbol{\alpha}'_\perp\mathbf{z}_t),$$

which shows that a system of N variables with k cointegrating relationships features $(N-k)$ linearly independent common trends (**TR**). The common trends are given by $(\boldsymbol{\alpha}'_\perp\mathbf{z}_t)$, while the coefficients on these trends are $\boldsymbol{\beta}_\perp(\boldsymbol{\alpha}'_\perp\boldsymbol{\beta}_\perp)^{-1}$. Note also that the stochastic trends depend on a set of initial conditions and cumulated disturbances:

$$\mathbf{TR}_t = \mathbf{TR}_{t-1} + \mathbf{C}(1)\,\mathbf{u}_t.$$

This brief discussion should have made it clear that the VECM and the MA models are complementary. As a consequence, the identification problem relevant for the vector of parameters in the cointegrating vectors $\boldsymbol{\beta}$ is also relevant for the vector of parameters determining the stochastic trends $\boldsymbol{\alpha}_{\perp}$. However, there is one aspect in which the two concepts are different. In theory, identified cointegrating relationships on a given set of variables should be robust to augmentation of the information set by adding new variables which should have a zero coefficient in the cointegrating vectors of the VECM representation of the larger information set. This is not true for the stochastic trends. Consider the case of augmenting an information set consisting of N variables admitting k cointegrating vectors to $N+M$ variables. The number of cointegrating vectors is constant while the number of stochastic trends increases by M; moreover, an unanticipated shock in the small system need not be unanticipated in the larger system. Note that we have added the qualifier "in theory" to our statement. In practice, given the size of available samples, the application of this procedure to analyse cointegration in a larger set of variables might lead to the identification of different cointegrating relationships from those obtained on a smaller set of variables.

6.2. *Forecasting from a Cointegrating VAR*

Assume a CVAR in N variables, of rank r, of the form

$$\Delta \mathbf{y}_t = \boldsymbol{\Pi}_0 + \boldsymbol{\Pi}_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{y}_{t-1} + \mathbf{u}_t,$$

or equivalently:

$$\Delta \mathbf{y}_t = \mathbf{g} + \boldsymbol{\Pi}_1 \left(\Delta \mathbf{y}_{t-1} - \mathbf{g}\right) + \boldsymbol{\alpha}\left(\boldsymbol{\beta}' \mathbf{y}_{t-1} - \mathbf{k}\right) + \mathbf{u}_t.$$

The system above can be re-parameterized as a first-order VAR:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \begin{bmatrix} \boldsymbol{\Pi}_0 \\ \boldsymbol{\beta}'\boldsymbol{\Pi}_0 \end{bmatrix} + \mathbf{v}_t$$

$$\mathbf{x}_t \equiv \begin{bmatrix} \Delta \mathbf{y}_t \\ \boldsymbol{\beta}' \mathbf{y}_t \end{bmatrix} \quad \mathbf{A} \equiv \begin{bmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\alpha} \\ \boldsymbol{\beta}'\boldsymbol{\Pi}_1 & I_r + \boldsymbol{\beta}'\boldsymbol{\alpha} \end{bmatrix} \quad \mathbf{v}_t \equiv \begin{bmatrix} \mathbf{u}_t \\ \boldsymbol{\beta}'\mathbf{u}_t \end{bmatrix}.$$

This VAR is stationary and hence it has a steady-state solution. By solving for this steady state and expressing variables in deviation form the steady state, we obtain:

$$\bar{\mathbf{x}}_t \equiv \begin{bmatrix} \Delta \mathbf{y}_t - \mathbf{g} \\ \boldsymbol{\beta}' \mathbf{y}_t - \mathbf{k} \end{bmatrix} = \mathbf{A}\bar{\mathbf{x}}_{t-1} + \mathbf{v}_t$$

$$\begin{bmatrix} \mathbf{g} \\ \mathbf{k} \end{bmatrix} = \left(\mathbf{I}_N - \mathbf{A}\right)^{-1} \begin{bmatrix} \boldsymbol{\Pi}_0 \\ \boldsymbol{\beta}'\boldsymbol{\Pi}_0 \end{bmatrix}.$$

At this point, it is possible to express the h-period ahead forecast as the cumulation of forecasts from the zero mean system:

$$
\begin{aligned}
E_t[\mathbf{y}_{t+h}] &= \mathbf{y}_t + \mathbf{g}h + \mathbf{J}\sum_{i=1}^{h}\mathbf{A}^i\bar{\mathbf{x}}_t \\
&= \mathbf{y}_t + \mathbf{g}h + \mathbf{B}_h\bar{\mathbf{x}}_t \\
&= \mathbf{y}_t + \mathbf{g}h + \boldsymbol{\alpha}_h\left(\boldsymbol{\beta}'\mathbf{y}_t - \mathbf{k}\right) + \boldsymbol{\Pi}_h\left(\Delta\mathbf{y}_t - \mathbf{g}\right) \\
\mathbf{B}_h &= \mathbf{J}\mathbf{A}\left(\mathbf{I}_{N+r} - \mathbf{A}\right)^{-1}\left(\mathbf{I}_{N+r} - \mathbf{A}^{h+1}\right)
\end{aligned}
$$

where $\mathbf{J} \equiv [\mathbf{I}_N,\ \mathbf{0}]$ is a selection matrix that picks $\Delta\mathbf{y}_t - \mathbf{g}$ out of the vector $\bar{\mathbf{x}}$. For h going to infinity, we obtain:

$$
\begin{aligned}
\lim_{h\to\infty} E_t\mathbf{y}_{t+h} - \mathbf{g}h &= \mathbf{y}_t + \mathbf{B}_\infty \\
&= \mathbf{y}_t + \boldsymbol{\alpha}_\infty\left(\boldsymbol{\beta}'\mathbf{y}_t - \mathbf{k}\right) + \boldsymbol{\Pi}_\infty\left(\Delta\mathbf{y}_t - \mathbf{g}\right).
\end{aligned}
$$

So the transitory components of the vector $\mathbf{y}_t$ are defined as:

$$
\mathbf{y}_t - \mathbf{y}_t^{BN} = -\boldsymbol{\alpha}_\infty\left(\boldsymbol{\beta}'\mathbf{y}_t - \mathbf{k}\right) - \boldsymbol{\Pi}_\infty\left(\Delta\mathbf{y}_t - \mathbf{g}\right).
$$

Note also that in the infinite horizon case, all disequilibria must be fully eliminated in expectations so the trend values must satisfy the cointegrating equilibrium:

$$
\boldsymbol{\beta}'\mathbf{y}_t^{BN} = \mathbf{k}.
$$

This condition obviously imposes restrictions on the parameters in $\boldsymbol{\alpha}_\infty$ and $\boldsymbol{\Pi}_\infty$.

6.3. *VECM and common trends representations*

The joint behaviour of stock prices and dividends under the dynamic dividend growth model is a good empirical example to illustrate VECM and common trend representations. Let's now decompose the (log) stock market prices lp_t in a permanent, information-related component, ld_t, and a temporary cyclical noise component v_t:

$$
\begin{aligned}
lp_t &= ld_t + v_t, \\
ld_t &= \mu_d + ld_{t-1} + u_t,
\end{aligned}
\tag{42}
$$

Dividends provide the stochastic trend in log-stock prices, which are the sum of a permanent component and of a transitory component; v_t and u_t are the shocks to the transitory and the permanent component of the system, respectively. These shocks are orthogonal and normally and independently distributed. Dividend and prices are cointegrated, in fact they share the single unobservable common stochastic trend in this system.

We obtain the VAR(1) representation by substituting for ld_t in the first equation from the second equation of (42):

$$
\begin{bmatrix} lp_t \\ ld_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_y \end{bmatrix} + \begin{bmatrix} 0\,1 \\ 0\,1 \end{bmatrix}\begin{bmatrix} lp_{t-1} \\ ld_{t-1} \end{bmatrix} + \begin{bmatrix} w_t \\ u_t \end{bmatrix},
$$

where $w_t = u_t + v_t$. From this model, we obtain the VECM representation:

$$\begin{bmatrix} \Delta lp_t \\ \Delta ld_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_y \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} lp_{t-1} \\ ld_{t-1} \end{bmatrix} + \begin{bmatrix} w_t \\ u_t \end{bmatrix},$$

where

$$\Pi = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} -1 & 1 \end{bmatrix} = \boldsymbol{\alpha}\boldsymbol{\beta}'.$$

The common trend representation is derived by considering that, because $lp_t - ld_t = v_t$, the MA representation for consumption and income growth is

$$\begin{bmatrix} \Delta lp_t \\ \Delta ld_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_y \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_t \\ u_t \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_{t-1} \\ u_{t-1} \end{bmatrix},$$

from which

$$\begin{bmatrix} lp_t \\ ld_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_y \end{bmatrix} t + \mathbf{C}^* (L) \begin{bmatrix} w_t \\ u_t \end{bmatrix} + C(1)\mathbf{z}_t,$$

where $\mathbf{z}_t$ is a process for which $\Delta \mathbf{z}_t = [w_t \ u_t]'$, and

$$\begin{aligned} \mathbf{C}(1) &= \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} = \boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}'_\perp \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}'_\perp, \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \left(\begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 & 1 \end{bmatrix}. \end{aligned}$$

Because in this application $\left(\boldsymbol{\alpha}'_\perp \boldsymbol{\beta}_\perp\right)^{-1} = 1$, dividends and prices have a single common stochastic trend. Such trend can be represented as

$$\boldsymbol{\alpha}'_\perp \left(\begin{bmatrix} \mu_y \\ \mu_y \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{t} w_t \\ \sum_{i=1}^{t} u_t \end{bmatrix} \right),$$

and only the shocks to the permanent component of prices enter the trend.

7. Risk, Returns and Portfolio Allocation with Cointegrated VARs

Consider the continuously compounded stock market return from time t to time $t+1$, $\mathbf{r}_{t+1}$. Define $\boldsymbol{\mu}_t$, the conditional expected log return given information up to time t, as follows

$$\mathbf{r}_{t+1} = \boldsymbol{\mu}_t + \mathbf{u}_{t+1},$$

where $\mathbf{u}_{t+1}$ is the unexpected log return. Define the k-period cumulative return from period $t+1$ through period $t+k$, as follows:

$$\mathbf{r}_{t,t+k} \equiv \sum_{i=1}^{k} \mathbf{r}_{t+i}.$$

The term structure of risk is defined as the conditional variance of cumulative returns, given the investor's information set, scaled by the investment horizon

$$\mathbf{\Sigma}_r(k) \equiv \frac{1}{k} Var[\mathbf{r}_{t,t+k}|D_t] \tag{43}$$

where $D_t \equiv \{\mathbf{z}_k\colon\ k \leq t\}$ consists of the full histories of returns as well as predictors that investors use in forecasting returns.

7.1. *Inspecting the mechanism: a bivariate case*

We illustrate the econometrics of the term structure of stock market risk by considering a simple bi-variate first-order VAR for continuously compounded total stock market returns, r_t^s, and the log dividend price,dp_t:

$$(\mathbf{z}_t - \mathbf{E}_z) = \mathbf{\Phi}_1 (\mathbf{z}_{t-1} - \mathbf{E}_z) + \boldsymbol{\nu}_t \qquad \boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\nu)$$

where

$$\mathbf{z}_t = \begin{bmatrix} r_t^s \\ dp_t \end{bmatrix} \quad \mathbf{E}_z = \begin{bmatrix} E_{r^s} \\ E_{d-p} \end{bmatrix} \quad \mathbf{\Phi}_1 = \begin{bmatrix} 0 \; \varphi_{1,2} \\ 0 \; \varphi_{2,2} \end{bmatrix}$$

$$\begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 \; \sigma_{12} \\ \sigma_{12} \; \sigma_2^2 \end{bmatrix} \right)$$

This bivariate model for returns and the predictor features a restricted dynamics such that only the lagged predictor is significant to determine current returns $(\varphi_{1,1} = 0)$ and the predictor is itself a strongly exogenous variable $(\varphi_{2,1} = 0)$. Given this VAR representation and the assumption of constant $\mathbf{\Sigma}_\nu$, one can see that

$$\begin{aligned} Var_t\left[(\mathbf{z}_{t+1} + ... + \mathbf{z}_{t+k})|D_t\right] = & \mathbf{\Sigma}_\nu + (\mathbf{I}_2 + \mathbf{\Phi}_1)\mathbf{\Sigma}_\nu(\mathbf{I}_2 + \mathbf{\Phi}_1)' + (\mathbf{I}_2 + \mathbf{\Phi}_1 + \mathbf{\Phi}_1^2)\mathbf{\Sigma}_\nu(\mathbf{I}_2 + \mathbf{\Phi}_1 + \mathbf{\Phi}_1^2)' + ... \\ & + (\mathbf{I}_2 + \mathbf{\Phi}_1 + ... + \mathbf{\Phi}_1^{k-1})\mathbf{\Sigma}_\nu(\mathbf{I}_2 + \mathbf{\Phi}_1 + ... + \mathbf{\Phi}_1^{k-1})', \end{aligned}$$

from which we derive:

$$\begin{aligned} \mathbf{\Sigma}_r(k) &= \frac{1}{k} \sum_{i=0}^{k-1} \mathbf{D}_i \mathbf{\Sigma} \mathbf{D}_i' \\ \mathbf{D}_i &= \mathbf{I}_2 + \mathbf{\Phi}_1 \mathbf{\Xi}_{i-1} \qquad i > 0 \\ \mathbf{\Xi}_i &= \mathbf{\Xi}_{i-1} + \mathbf{\Phi}_1^i \qquad i > 0 \\ \mathbf{D}_0 &\equiv \mathbf{I}_2, \quad \mathbf{\Xi}_0 \equiv \mathbf{I}_2 \end{aligned}$$

Note that, under the chosen specification of the matrix $\mathbf{\Phi}_1$ we can write the generic term

$\mathbf{D}_i\mathbf{\Sigma}\mathbf{D}_i'$, as follows:

$$\mathbf{D}_i\mathbf{\Sigma}\mathbf{D}_i' = \left(\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{12}' & M_{22} \end{array}\right) \tag{44}$$

$$\begin{aligned} M_{11} &= \Sigma_{1,1} + \Phi_{1,2}\Xi_{i-1}^{(22)}\Sigma_{1,2}' + \Sigma_{1,2}\Xi_{i-1}^{(22)\prime}\Phi_{1,2}' + \Phi_{1,2}\Xi_{i-1}^{(22)}\Sigma_{2,2}\Xi_{i-1}^{(22)\prime}\Phi_{1,2}' \\ M_{12}' &= \Xi_i^{(22)}\Sigma_{1,2}' + \Xi_i^{(22)}\Sigma_{2,2}\Xi_{i-1}^{(22)\prime}\Phi_{1,2}' \\ M_{22} &= \Xi_i^{(22)}\Sigma_{2,2}\Xi_i^{(22)\prime} \end{aligned}$$

where we have used the fact that

$$\mathbf{\Xi}_i = \sum_{j=0}^{i} \mathbf{\Phi}_1^i = \left(\begin{array}{c|c} 0 & \phi_{1,2}\sum_{j=0}^{i-1}\phi_{2,2}^j \\ \hline 0 & \sum_{j=0}^{i}\phi_{2,2}^j \end{array}\right) \quad D_i = I + \Phi_1\Xi_{i-1} = \left(\begin{array}{c|c} I & \phi_{1,2}\sum_{j=0}^{i-1}\phi_{2,2}^j \\ \hline 0 & \sum_{j=0}^{i}\phi_{2,2}^j \end{array}\right).$$

Equation ((44)) implies that, in our simple bivariate example, the term structure of stock market risk takes the form

$$\sigma_r^2(k) = \sigma_1^2 + 2\varphi_{1,2}\sigma_{1,2}\psi_1(k) + \varphi_{1,2}^2\sigma_{2,2}^2\psi_2(k) \tag{45}$$

where

$$\begin{aligned} \psi_1(k) &\equiv \frac{1}{k}\sum_{l=0}^{k-2}\sum_{i=0}^{l}\varphi_{2,2}^i \quad k > 1 \\ \psi_2(k) &\equiv \frac{1}{k}\sum_{l=0}^{k-2}\left(\sum_{i=0}^{l}\varphi_{2,2}^i\right)^2 \quad k > 1 \\ \psi_1(1) &= \psi_2(1) = 0. \end{aligned}$$

Total stock market risk can be decomposed in three components: IID uncertainty, measured by σ_1^2, mean reversion, $2\varphi_{1,2}\sigma_{1,2}\psi_1(k)$, and uncertainty about future predictors, $\varphi_{1,2}^2\sigma_{2,2}^2\psi_2(k)$. Without predictability $\left(\varphi_{1,2} = 0\right)$ the entire term structure is flat at the level σ_1^2. This is the classical situation where portfolio choice is independent of the investment horizon. The possible downward slope of the term structure of risk depends on the second term, and it is therefore crucially affected by predictability and a negative correlation between the innovations in dividend price ratio and in stock market returns $(\sigma_{1,2})$; the third term is always positive and increasing with the horizon when the autoregressive coefficient in the dividend yield process is positive. Overall, the slope of the term structure of risk depends on the significance of the dividend-price in explaining returns, on the contemporaneous correlation between the innovations in the equations for the dividend-price and returns, on the variance of returns and the dividend-price, and on the persistence of the dividend-price.

Table 1: A simple bivariate VAR (1910-2008)

$$\left(r^s_{t+1} - E_{r^s}\right) = \varphi_{12}\left(dp_t - E_{dp}\right) + \upsilon_{1t+1}$$
$$\left(dp_{t+1} - E_{dp}\right) = \varphi_{22}\left(dp_t - E_{dp}\right) + \nu_{2t+1}$$

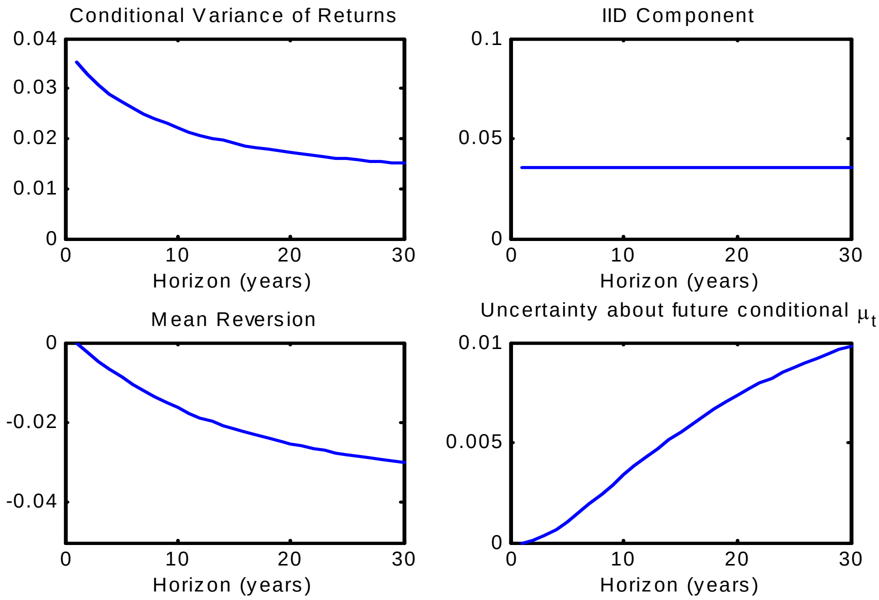
$\underset{(t-stat)}{\varphi_{12}}$	$\underset{(t-stat)}{\varphi_{22}}$	$\underset{\varphi_{11}=0,\varphi_{21}=0}{\chi^2_2}$	σ_1	σ_2	$\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}$	$adjR^2_{r^s_{t+1}}$	$adjR^2_{dp_{t+1}}$
0.073	0.893	3.128	0.196	0.208	-0.844	0.02	0.79
(1.71)	(19.70)	(0.21)					

Table 1: The table reports coefficient estimates (with t-statistics in parentheses) and the R^2 statistic for each equation. We also report the standard deviations and correlations of residuals.

Table 1 summarizes the results of the estimation of the system.[6] The estimation results confirm the noisy nature of 1-year stock market returns and the high persistence of the dividend-price ratio. The covariance structure of the innovations is such that the unexpected log excess stock returns are highly negatively correlated with the innovations in the log dividend price ratio. The following Figure plots the term structure of risk resulting from the estimation of the restricted VAR and its decomposition. The evidence of a downward sloping curve with risk halving when moving from a one-year to a thirty-year horizon replicates the results in Campbell and Viceira (2002), based on the estimation of a larger model including bond and stock excess returns, the nominal and real risk free rate together with the dividend-

[6] We consider a dataset of annual observations for the period 1910-2009. The data are from Welch and Goyal (2008), who provide detailed descriptions of the data and their sources. Stock returns are measured as continuously compounded returns on the S&P 500 index, including dividends. To compute real returns we calculate the inflation rate from the CPI (for all urban consumers). The predictor for the equity premium is the dividend-price ratio, computed as the difference between the log of dividends paid on the S&P 500 index and log of stock prices (S&P 500 index), where dividends are measured using a one-year moving sum.

yield and the yield spread as predictors.



The TS of stock market risk from a bi-variate VAR

7.2. *A VAR with many assets and predictors*

Consider now a more articulated model with many asset classes and predictors. Following Barberis (2000) and Campbell and Viceira (2002, 2005), we describe asset return dynamics by means of a first-order vector autoregressive or VAR(1) model. We choose a VAR(1) as the inclusion of additional lags, even if feasible, may reduce the precision of the estimates:

$$\mathbf{z}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{z}_{t-1} + \boldsymbol{\nu}_t, \tag{46}$$

where $\mathbf{z}_t \equiv [r_{0t}\ \mathbf{x}_t\ s_t]'$ is a $M \times 1$ vector, with r_{0t} being the log real return on the asset used as a benchmark to compute excess returns on all other asset classes, $\mathbf{x}_t$ being the $M \times 1$ vector of log excess returns on all asset classes with respect to to the benchmark, and $\mathbf{s}_t$ is the $(M - N - 1) \times 1$ vector of return predictors. In the VAR(1) specification, $\mathbf{\Phi}_0$ is a $m \times 1$ vector of intercepts and $\mathbf{\Phi}_1$ is a $M \times M$ matrix of slopes. Finally, $\boldsymbol{\nu}_t$ is a $M \times 1$ vector of innovations in asset returns and return predictors, for which standard assumptions apply, i.e.:

$$\boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\nu) \tag{47}$$

where $\mathbf{\Sigma}_\nu$ is the $M \times M$ covariance matrix. Note that

$$\mathbf{\Sigma}_\nu = \begin{bmatrix} \sigma_0^2 & \boldsymbol{\sigma}'_{0x} & \boldsymbol{\sigma}'_{0s} \\ \boldsymbol{\sigma}_{0x} & \mathbf{\Sigma}_{xx} & \mathbf{\Sigma}'_{xs} \\ \boldsymbol{\sigma}_{0s} & \mathbf{\Sigma}_{xs} & \mathbf{\Sigma}_{ss} \end{bmatrix}$$

and the unconditional mean and covariance matrix of $\mathbf{z}_t$, assuming that the VAR is stationary and therefore that its moments are well-defined and time-invariant, can be represented as follows:

$$\boldsymbol{\mu}_z = (\mathbf{I}_M - \boldsymbol{\Phi}_1)^{-1}\, \boldsymbol{\Phi}_0$$
$$vec\,(\boldsymbol{\Sigma}_{zz}) = (\mathbf{I}_{M^2} - \boldsymbol{\Phi}_1 \otimes \boldsymbol{\Phi}_1)^{-1}\, vec\,(\boldsymbol{\Sigma}_{\nu})\,.$$

The conditional mean and variance of the cumulative asset returns at different horizons are instead:

$$E_t[\mathbf{z}_{t+1} + ... + \mathbf{z}_{t+K}] = \left(\sum_{i=0}^{k-1} (k-i)\, \boldsymbol{\Phi}_1^i\right) \boldsymbol{\Phi}_0 + \left(\sum_{j=0}^{k} \boldsymbol{\Phi}_1^j\right) \mathbf{z}_t$$

$$Var_t[\mathbf{z}_{t+1} + ... + \mathbf{z}_{t+K}] = \boldsymbol{\Sigma}_{\nu} + (\mathbf{I}_M + \boldsymbol{\Phi}_1)\boldsymbol{\Sigma}_{\nu}(\mathbf{I}_M + \boldsymbol{\Phi}_1)' + (\mathbf{I}_M + \boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_1^2)\boldsymbol{\Sigma}_{\nu}(\mathbf{I}_M + \boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_1^2)' + ...$$
$$+ (\mathbf{I}_M + \boldsymbol{\Phi}_1 + ... + \boldsymbol{\Phi}_1^{K-1})\boldsymbol{\Sigma}_{\nu}(\mathbf{I}_M + \boldsymbol{\Phi}_1 + ... + \boldsymbol{\Phi}_1^{K-1})'$$

Once the conditional moments of excess returns are available, the following selector matrix extracts for each period, k-period conditional moments of log real returns:

$$\mathbf{M}_r = \begin{bmatrix} 1 & \mathbf{0}'_N & \mathbf{0}'_{M-N-1} \\ \boldsymbol{\iota}_N & \mathbf{I}_N & \mathbf{O}_{N\times(M-N-1)} \end{bmatrix}$$

which implies

$$\frac{1}{k}\begin{bmatrix} E_t\left(r_{0,t+1}^k\right) \\ E_t\left(\mathbf{r}_{t+1}^k\right) \end{bmatrix} = \frac{1}{k}\mathbf{M}_r E_t[\mathbf{z}_{t+1} + ... + \mathbf{z}_{t+K}]$$
$$\frac{1}{k}\begin{bmatrix} Var_t\left(r_{0,t+1}^k\right) \\ Var_t\left(\mathbf{r}_{t+1}^k\right) \end{bmatrix} = \frac{1}{k}\mathbf{M}_r Var_t[\mathbf{z}_{t+1} + ... + \mathbf{z}_{t+K}]\mathbf{M}'_r$$

Therefore after the estimation of the VAR, it is possible to derive unconditional and conditional moments for returns and excess returns at all investment horizons. These moments deliver the dynamics of returns and the risk of different assets across investment horizons. This information forms the input to portfolio allocation.

7.3. *Mean-variance analysis with a VAR model*

The starting point of mean-variance analysis is an expression for the log-returns on the portfolio. Campbell and Viceira (1999) show that the log-return on the portfolio can be approximated as follows:

$$r_{p,t+1} = r_{0,t+1} + \boldsymbol{\alpha}'_t \mathbf{x}_t + \frac{1}{2}\boldsymbol{\alpha}'_t\left(\boldsymbol{\sigma}_x^2 - \boldsymbol{\Sigma}_{xx}\boldsymbol{\alpha}_t\right)$$
$$\mathbf{x}_t = (\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}_N)$$
$$\boldsymbol{\Sigma}_{xx} = Var_t\left(\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}_N\right) \qquad \boldsymbol{\sigma}_x^2 = diag\left(\boldsymbol{\Sigma}_{xx}\right).$$

Given these definitions different problems can be addressed. First, Campbell and Viceira (2004) show that the optimal weights $\omega_{T,t}$ for the tangency portfolio (i.e., the portfolio that with no loading on the risk -free asset) in the one-period mean-variance frontier are obtained by considering the risk-free asset as a benchmark in the VAR and by using the following closed-form expression that is obtained from a standard first-order condition:

$$\boldsymbol{\omega}_{T,t} = \lambda_f \boldsymbol{\Sigma}_{xx}^{-1} \left[E_t \left(\mathbf{r}_{t+1} - r_{0,t+1} \boldsymbol{\iota}_N \right) + \frac{1}{2} \boldsymbol{\sigma}_x^2 \right]$$

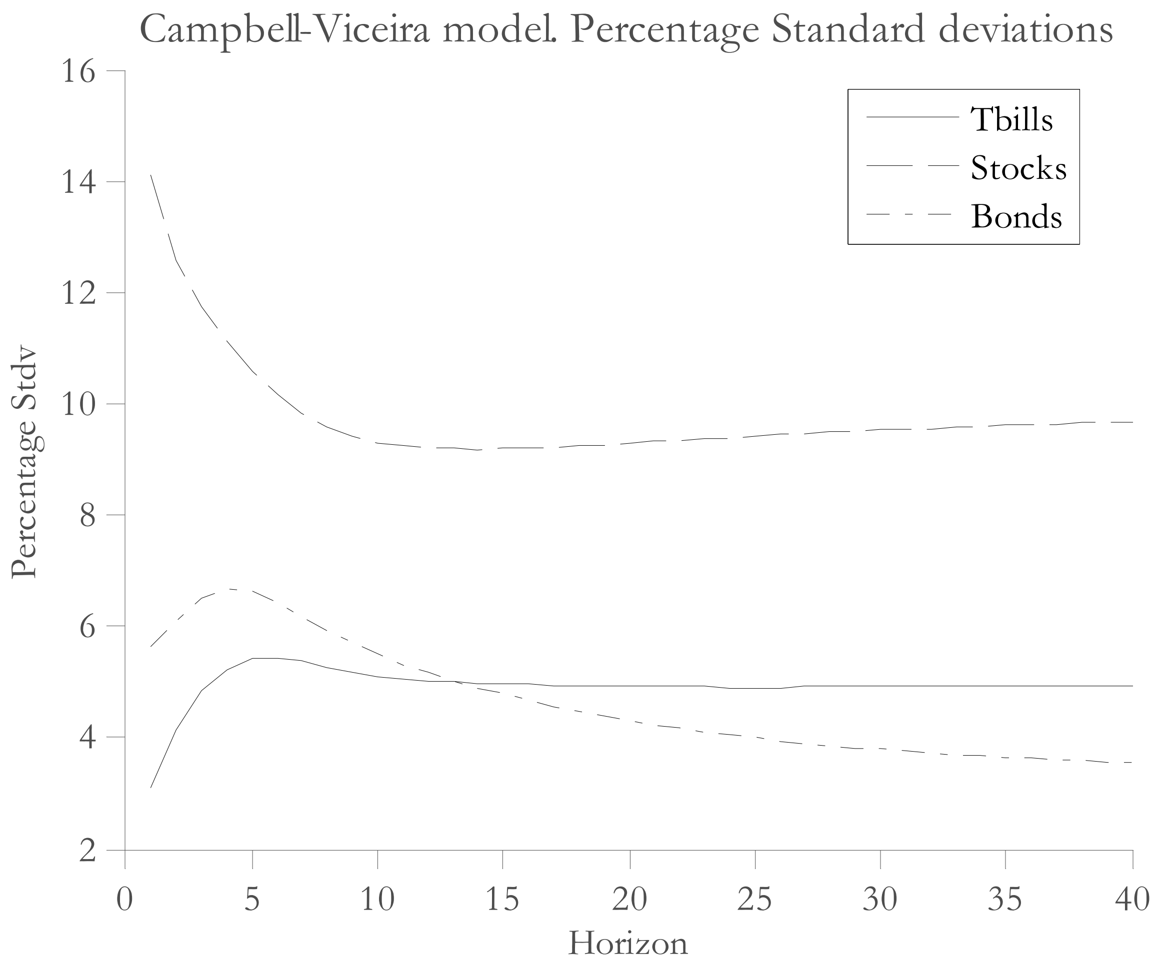
$$\lambda_f \equiv \frac{1}{\left[E_t \left(\mathbf{r}_{t+1} - r_{0,t+1} \boldsymbol{\iota}_N \right) + \frac{1}{2} \boldsymbol{\sigma}_x^2 \right]' \left(\boldsymbol{\Sigma}_{xx}^{-1} \right)' \boldsymbol{\iota}_N}.$$

Considering a k-period horizon we have instead:

$$\omega_{T,t}(k) = \lambda_f(k) \boldsymbol{\Sigma}_{xx}^{-1}(k) \left[E_t(\mathbf{r}_{t+1}^{(k)} - r_{0,t+1}^{(k)} k \boldsymbol{\iota}_N) + \frac{1}{2} \boldsymbol{\sigma}_x^2(k) \right]$$

$$\lambda_f(k) \equiv \frac{1}{\left[E_t(\mathbf{r}_{t+1}^{(k)} - r_{0,t+1}^{(k)} k \boldsymbol{\iota}_N) + \frac{1}{2} \boldsymbol{\sigma}_x^2(k) \right]' \left(\boldsymbol{\Sigma}_{xx}^{-1}(k) \right)' \boldsymbol{\iota}_N}$$

This way, a VAR can be put at work to derive optimal portfolio weights when expected returns are set equal to their long-term sample means but variance-covariance of returns change across investment horizons according to VAR estimates. The typical empirical evidence produced by VAR models is the following term structure of risk:



References

[1] Barberis, N., (2000), "Investing for the long run when returns are predictable ", *Journal of Finance*, vol.55, no.1, pages 225-264.

[2] Bec, F., and C. Gollier, (2008), "Asset returns volatility and investment horizon: The French case", *working paper*.

[3] Campbell J., (1999), "Asset prices, consumption, and the business cycle", *Handbook of Macroeconomics*, vol.1, pages 1231-1303.

[4] Campbel J., A. Lo and A. MacKinley, (1997), *The Econometrics of Financial Markets*, Princeton University Press.

[5] Campbell, J., and L. Viceira, (2002), *Strategic Asset Allocation*, Oxford University Press.

[6] Campbell, J., and L. Viceira, (2005), "The term structure of risk-return trade-off", *NBER working paper*, no. 11119.

[7] Merton, R.C., (1969), "Lifetime portfolio selection under uncertainty: the continuous-time case", *Review of Economics and Statistics*, vol. 51, no. 3, pages 247-257.

[8] Merton, R.C., (1973), "An intertemporal capital asset pricing model", *Econometrica*, vol. 41, no. 5, pages 867-887.

[9] Samuelson, P., (1969) "Lifetime portfolio selection by dynamic stochastic programming ", *Review of Economics and Statistics*, vol.51, no.3, pages 239-246.

[10] Stambaugh, R.F., (1999), "Predictive regressions", *Journal of Financial Economics*, vol. 53, no. 3, pages, 375-421.

[11] Ang, Andrew and Geert Bekaert, 2007, Stock Return Predictability: Is It There? The Review of Financial Studies, 20, 651–707.

[12] Ang, A. and A. Maddaloni."Do Demographic Changes Affect Risk Premiums? Evidence from International Data." *Journal of Business*, 78 (2005), 341-380.

[13] Bakshi, Gurdip S., and Zhiwu Chen, 1994, Baby Boom, Population Aging, and Capital Markets, Journal of Business, 67, 2, 165-202.

[14] Van Binsbergen J.H. and R.S.Koijen, 2009, Predictive Regression:a Present Value Approach, mimeo

[15] Boudoukh, Jacob, Richardson, Matthew and Robert F. Whitelaw, 2008, The Myth of Long-Horizon Predictability, The Review of Financial Studies, 21, 4, 1577-1605.

[16] Campbell, J. Y., 1991. A variance decomposition for stock returns. Economic Journal 101, 157–179.

[17] Campbell, J. Y., Lo, A. W., MacKinlay, C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, NJ.

[18] Campbell, John Y., and Samuel B. Thomson, 2008, Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, The Review of Financial Studies, 21, 1509-1531.

[19] Campbell, John Y., and Robert Shiller, 1988, Stock Prices,Earnings, and Expected

Dividends, Journal of Finance, 43, 661–676.

[20] Campbell, John Y., and Luis M. Viceira, 2002, Strategic Asset Allocation: Portfolio Choice for Long-Term Investors (Oxford University Press).

[21] Campbell, John Y., and Luis M. Viceira, 2005, The Term Structure of the Risk-Return Tradeoff, Financial Analyst Journal, 61, 1, 34-44

[22] Cochrane, J.H., 1994 Permanent and Transitory Components in GNP and Stock Prices, Quarterly Journal of Economics

[23] Cochrane, J. H., 2001. Asset Pricing. Princeton University Press, Princeton, NJ.

[24] Cochrane, John H., 2008, The Dog that Did Not Bark: A Defense of Return Predictability, Review of Financial Studies, 20, 5.

[25] Cochrane J.H., 2008, State-Space versus VAR models of Stock Returns, mimeo

[26] DellaVigna S., and J.Pollet. "Demographics and Industry Returns." *American Economic Review*, 97(2007), 1167-1702.

[27] Erb, C. B.; C. R. Harvey, and T. E. Viskanta. "Demographics and International Investment." *Financial Analysts Journal*, 53 (1996), 4,14-28.

[28] Fama, Eugene and Kenneth R. French, 1988, Dividend Yields and Expected Stock Returns, Journal of Financial Economics, 22, 3-26.

[29] Favero C.A., A.Gozluklu and A.Tamoni (2010) "Demographic Trends, the Dividend-Price Ratio and the Predictability of Long-Run Stock Market Returns", CEPR working paper 7734, forthcoming in the Journal of Financial and Quantitative Analysis

[30] Geanakoplos, John, Magill, Michael and Martine Quinzii, 2004, Demography and the Long Run Behavior of the Stock Market, Brookings Papers on Economic Activities, 1: 241-325.

[31] Goyal, A., Welch, I., May 2003. Predicting the equity premium with dividend ratios. Management Science 49 (5), 639–654.

[32] Goyal, A. "Demographics, Stock Market Flows, and Stock Returns." *Journal of Financial and Quantitative Analysis*, 39 (2004), 1, 115-142.

[33] Goyal, Amit, and Ivo Welch, 2008, A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. The Review of Financial Studies, 21-4, 1455-1508

[34] Hodrick, Robert, 1992, Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, Review of Financial Studies, 5, 357–386.

[35] Lacerda, F., and Santa-Clara, P. 2010, Forecasting Dividend Growth to Better Predict Returns, Discussion paper.

[36] Lettau, Martin, and Sydney Ludvigson, 2005, Expected Returns and Expected Dividend Growth, Journal of Financial Economics, 76, 583-626.

[37] Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, Aggregate Wealth and Expected Stock Returns, Journal of Finance, 56, 3, 815-849.

[38] Lettau, Martin, and Stijn Van Nieuwerburgh, 2008, Reconciling the Return Predictability Evidence, Review of Financial Studies, 21, 4, 1607-1652.

[39] Lewellen, Jonathan, 2004, Predicting Returns with Financial Ratios. Journal of Financial Economics 74:209–35.

[40] Nelson, C. C., Kim, M. J., 1993. Predictable stock returns: The role of small sample bias. Journal of Finance 43, 641–661.

[41] Neely C.J., Weller P.A.(2000). Predictability in international asset returns: a reexamination. Journal of Financial and Quantitative Analysis, 35, 601-620

[42] Newey, Whitney K. and Kenneth D. West, 1987, A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, 55, 3, 703-08.

[43] Newey, Whitney K. and Kenneth D. West, 1994, Automatic Lag Selection in Covariance Matrix Estimation, Review of Economic Studies, 61, 631-653.

[44] Paye, B.S., Timmermann A.(2006), Instability of return prediction models. Journal oF Empirical Finance, 13, 274-315

[45] Pesaran M.H., A.Pick, and A.G.Timmermann(2010), Variable Selection Estimation and Inference for Multi-period forecasting problems, CEPR working paper 7139

[46] Poterba, J. M. "Demographic Structure and Asset Returns." *The Review of Economics and Statistics*, 83 (2001), 4, 565-584.

[47] Rapach D.E., Wohar M.E., 2006, In-sample vs. out-of-sample tests of stock return predictability in the context of data-mining, Journal of Empirical Finance, 13, 231-247

[48] Robertson D. and S.Wright, 2009, The Limits to Stock Return Predictability, mimeo

[49] Oleg Rytchkov, 2008, Filtering Out Expected Dividends and Expected Returns, Working paper.

[50] Siegel Jeremy.J.(1998), Stocks For the Long-Run, 2nd edition, McGraw-Hill

[51] Shiller, Robert J., 2005, Irrational Exuberance, second edition, Princeton University Press.

[52] Schotman, Peter C., Rolf Tschernig, and Jan Budek, Long Memory and the Term Structure of Risk, Journal of Financial Econometrics, Fall 2008, 6 (4), 459495.

[53] Stambaugh, R. F., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.

[54] Taleb, N.N., 2001, Fooled By Randomness. The Hidden Role of Chance in the Markets and in Life, TEXERE XLC

[55] Valkanov, R., 2003. Long-horizon regressions: Theoretical results and applications. Journal of Financial Economics 68, 201–232. 33