Table of Contents

K-C Chiu & L Xu, (2003) "White noise tests and synthesis of APT economic factors using TFA", *Computational Intelligence in Economics and Finance*, SH Chen, P Wang (Ed.): Series on Advanced Information Processing (series editor: L. Jain), pp 405-419, Springer Verlag.

1. White Noise Tests and Synthesis of APT Economic Factors Using TFA

Kai Chun Chiu and Lei Xu

Department of Computer Science and Engineering,
The Chinese University of Hong Kong,
Shatin, N.T., Hong Kong, P.R. China.
Email: {kcchiu,lxu}@cse.cuhk.edu.hk

When the Temporal Factor Analysis (TFA) model is used for classical Arbitrage Pricing Theory (APT) analysis in finance, it is necessary to perform white noise tests on the residual for model adequacy. We carry out white noise tests and obtain results that provide assurance for further statistical analysis using the TFA model. We also explore empirically the relationship between macroeconomic time series and Gaussian statistically uncorrelated hidden factors recovered by TFA. Based on the statistical hypothesis test results, we conclude that each of the four economic time series is linearly related to the uncorrelated factors. Consequently, APT economic factors can be synthesized from those statistically uncorrelated factors.

1.1 Introduction

Since its origination by Ross [1.10] in 1976, the Arbitrage Pricing Theory (APT) has drawn the attention of the finance community worldwide. APT relates security returns to a spectrum of several risk factors. These are often called economic factors because in literature time series of macroeconomic variables have been found probable candidates to be used as proxies for the purpose of financial modelling.

Although many arguments have been put forward for the selection of proper proxies, there remains no consensus on what should be the most appropriate proxies for those hidden economic factors over which APT seeks to model. For instance, Estep, Hanson and Johnson have considered possible factors to be changes in inflation, real growth, oil prices, defense spending and real interest rates in [1.3] in 1983. On the other hand, in 1984 Roll and Ross assert that an asset's return being directly related to unanticipated changes in four economic variables and they are inflation, industrial production, risk premiums and the slope of the term structure of interest rates respectively [1.9]. Interestingly, Chen, Roll and Ross in their paper [1.2] two years later change the last two factors to be the spread between long and short interest rates and the spread between high- and low-grade bonds. Furthermore, it has been proposed in [1.1] in 1988 that the factors can possibly be usually spread between the total monthly returns on government and corporate bonds and Treasury bills, unexpected deflation and growth rate in real final sales.

The lack of consensus and consistency over what should be the real economic factors in APT can largely be ascribed to the lack of systematic and effective methods for recovering those hidden economic factors. However, a new factor analytic technique termed Temporal Factor Analysis (TFA) proposed in [1.14] has been found to be capable of extracting statistically uncorrelated Gaussian factor scores from time series of stationary stock returns. Moreover, it provides a reasonable basis for the synthesis of stationary time series of economic variables.

The rest of this paper will be divided into five sections. Section 2 briefly reviews the arbitrage pricing theory and section 3 gives an overview of the TFA model. Statistical tests for serial correlations of the residual components for model adequacy will be presented in section 4, which is followed by the analysis and results of economic factors synthesis in section 5. Section 6 would be devoted to concluding remarks.

1.2 The Arbitrage Pricing Theory

The APT begins with the assumption that the $n \times 1$ vector of asset returns, R_t, is generated by a linear stochastic process with k factors [1.10, 1.9, 1.8]:

$$R_t = \bar{R} + Af_t + e_t \tag{1.1}$$

where f_t is the $k \times 1$ vector of realizations of k common factors, A is the $n \times k$ matrix of factor weights or loadings, and e_t is a $n \times 1$ vector of asset-specific risks. It is assumed that f_t and e_t have zero expected values so that $\bar{R}$ is the $n \times 1$ vector of mean returns. The model addresses how expected returns behave in a market with no arbitrage opportunities and predicts that an asset's expected return is linearly related to the factor loadings or

$$\bar{R} = R_f + Ap \tag{1.2}$$

where R_f is a $n \times 1$ vector of constants representing the risk-free return, and p is $k \times 1$ vector of risk premiums. Similar to the derivation of CAPM, (1.2) is based on the rationale that unsystematic risk is diversifiable and therefore should have a zero price in the market with no arbitrage opportunities.

1.3 Temporal Factor Analysis

1.3.1 An Overview of TFA

Suppose the relationship between a state $y_t \in \mathrm{R}^k$ and an observation $x_t \in \mathrm{R}^d$ are described by the first-order state-space equations as follows [1.14, 1.13]:

$$y_t = By_{t-1} + \varepsilon_t, \tag{1.3a}$$
$$x_t = Ay_t + e_t, \qquad t = 1, 2, \ldots, N. \tag{1.3b}$$

where ε_t and e_t are mutually independent zero-mean white noises with $E(\varepsilon_i \varepsilon_j) = \Sigma_\varepsilon \delta_{ij}$, $E(e_i e_j) = \Sigma_e \delta_{ij}$, $E(\varepsilon_i e_j) = 0$, Σ_ε and Σ_e are diagonal matrices, and δ_{ij} is the Kronecker delta function:

$$\delta_{ij} = \begin{cases} 1, \text{if } i = j, \\ 0, \text{otherwise.} \end{cases} \tag{1.4}$$

We call ε_t driving noise upon the fact that it drives the source process over time. Similarly, e_t is called measurement noise because it happens to be there during measurement. The above model is generally referred to as the TFA model.

In the context of APT analysis, (1.1) can be obtained from (1.3b) by substituting $(\tilde{R}_t - \bar{R})$ for x_t and f_t for y_t. The only difference between the APT model and the TFA model is the added (1.3a) for modelling temporal relation of each factor. The added equation represents the factor series $y = \{y_t\}_{t=1}^T$ in a multi-channel auto-regressive process, driven by an i.i.d. noise series $\{\varepsilon_t\}_{t=1}^T$ that are independent of both y_{t-1} and e_t. Specifically, it is assumed that ε_t is Gaussian distributed. Moreover, TFA is defined such that the k sources $y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(k)}$ in this state-space model are statistically independent. This constraint implies B is diagonal and ε_t is mutually independent in components. The objective of TFA [1.13, 1.14] is to estimate the sequence of y_t's with unknown model parameters $\Theta = \{A, B, \Sigma_\varepsilon, \Sigma_e\}$ through available observations. Since for Gaussian distribution statistically independent is synonymous with uncorrelated, we will use the two terms interchangeably in this paper.

In implementation, an adaptive algorithm has been suggested. At each time unit, factor loadings are estimated by cross-sectional regression and factor scores are estimated by maximum likelihood learning. Here, we adopt a simple algorithm proposed in [1.14] as shown below.

Assume $G(\varepsilon_t|0, I)$ and $G(e_t|0, \Sigma)$.

- **Step 1** Fix A, B and Σ, estimate the hidden factors y_t by

$$\hat{y}_t = [I + A^T \Sigma^{-1} A]^{-1} (A^T \Sigma^{-1} \bar{x}_t + B\hat{y}_{t-1}), \tag{1.5}$$
$$\varepsilon_t = \hat{y}_t - B\hat{y}_{t-1}, \tag{1.6}$$
$$e_t = \bar{x}_t - A\hat{y}_t, \tag{1.7}$$

- **Step 2** Fix, $\hat{y}_t$, update B, A and Σ_e by gradient descent method as follows:

$$B^{new} = B^{old} + \eta \text{diag}[\varepsilon_t \hat{y}_{t-1}^T], \tag{1.8}$$
$$A^{new} = A^{old} + \eta e_t \hat{y}_t^T, \tag{1.9}$$
$$\Sigma^{new} = (1 - \eta)\Sigma^{old} + \eta e_t e_t^T. \tag{1.10}$$

1.3.2 Model Selection vs Appropriate Number of Factors

Central to the discussion in the paper about the number of factors in APT, TFA is superior to maximum likelihood factor analysis (MLFA) in view of its model selection ability. In the context of APT analysis, the scale or complexity of the model is equivalent to the number of hidden factors in the original factor structure. As a result, model selection refers to deciding the appropriate number of factors in APT. We can achieve the aim of model selection by enumerating the cost function $J(k)$ with k incrementally and then select an appropriate k by [1.14, 1.12, 1.15]

$$\min_k J(k) = \frac{1}{2}[k \ln(2\pi) + k + \ln |\Sigma|] \tag{1.11}$$

1.3.3 Grounds and Benefits for Using TFA in APT Analysis

Firstly, it is assumed that factors follow Gaussian distribution at each time t. There is a consensus that the noisy component in most econometric and statistical models being Gaussian distributed. The rationale comes from the central limit theorem which implies that the compounding of a large number of unknown distributions will be approximately normal. Secondly, we believe that factors recovered must be independent of each other. Although economic factors are seldom independent, it is helpful to discover statistically independent factors for the purpose of analysis because the restriction of independence will rule out many possible solutions which contain redundant elements. Furthermore, economic interpretation of factors recovered can be easily achieved by appropriate combination of those independent factors. Thirdly, we believe there exists temporal relation between factors. Equation (1.3a) models an AR(1) time series of factors. Although formulation in this way slightly deviates from the Efficient Market Hypothesis (EMH) [1.4] on which APT is based, its rationale can be found in the literature [1.7, 1.5] which describes empirical evidence against the EMH.

Compared with MLFA, TFA has at least three distinct benefits. First, with the independence assumption in the derivation, the recovered factors are assured to be statistically independent. Second, Xu in [1.13] has shown that taking into account temporal relation effectively removes rotation indeterminacy. As a result, the solution given by TFA is unique. Theorem 3 proved by Xu in [1.13] illustrates this point. Third, it can determine the number of hidden factors via its model selection ability. Furthermore, it should be noted that MLFA is a special case of the model with $B = 0$ in (1.3a).

1.3.4 Testability of the TFA Model

The TFA model retains virtually all statistical properties of the original APT model. It is simply an extension of the APT model because it additionally includes temporal relation between factors in the APT model. Apart from that, there is no significant difference. Since the relationship between y_t and y_{t-1} described by the added equation is also linear, the entire TFA model is a linear model with both the driving noise ε_t and the measurement noise e_t assumed to be Gaussian distributed. Moreover, as both the returns and factors are stationary and the factors is assumed to be uncorrelated with idiosyncratic risks, we can safely conclude that the model is testable, just like APT [1.8, 1.11, 1.6].

1.4 Tests of White Noise Residuals

It is common in the literature of statistics that tests for model adequacy should immediately follow parameter estimation of the model under consideration. Usually the model is considered adequate if the residual component consists of white noise. In section 1.3 we have specifically emphasized the residual components of the TFA model. They are ε_t, the driving noise in (1.3a) and e_t, the measurement noise in (1.3b) respectively. For the AR(1) model as given by (1.3a) to be adequate, the estimated driving noise component should be substantially serially uncorrelated, i.e., autocorrelation of its lags should not be significantly different from zero. At the same time, for (1.3b) to be adequate, the estimated observation noise component should be largely uncorrelated among its constituents.

1.4.1 Data Considerations

We have carried out our analysis using past stock price and return data of Hong Kong. Daily closing prices of 86 actively trading stocks covering the period from January 1, 1998 to December 31, 1999 are used. The number of trading days throughout this period is 522. These stocks can be subdivided into three main categories according to different indices they constitute. Of the 86 equities, 30 of them belongs to the Hang Seng Index (HSI) constituents, 32 are Hang Seng China-Affiliated Corporations Index (HSCCI) constituents and the remaining 24 are Hang Seng China Enterprises Index (HSCEI) constituents. Before carrying out the analysis, the stock prices have been converted to stationary stock returns via $R_t = \frac{p_t - p_{t-1}}{p_{t-1}}$.

1.4.2 Test Statistics

To check if the driving noise residuals behave as a white-noise process, we will adopt the Ljung-Box modified Q-statistic shown below. The Q-statistic can be used to test whether a group of autocorrelations is significantly different from zero.

$$Q = N(N+2)\sum_{k=1}^{s}\frac{r_k^2(\hat{\varepsilon})}{N-k} \tag{1.12}$$

where N is the effective number of observations and s is the lag order. If the sample value of Q calculated above exceeds the critical value of χ^2 with $s-1$ degrees of freedom at $\alpha = 5\%$, then we can conclude that at least one value of r_k is statistically different from zero at 5% level of significance and suspect the residuals are serially correlated and not white.

On the other hand, to investigate whether each cross correlation coefficient of the observation noise residuals is not significantly different from zero, we will apply the t-test with the test statistic given by

$$t = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}, \tag{1.13}$$

where r is the correlation coefficient of a sample of n points (x_i, y_i) as given by

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{[\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2]^{\frac{1}{2}}}. \tag{1.14}$$

Assume that the x and y values originate from a bivariate Gaussian distribution, and that the relationship is linear, it can be shown that t follows Student's t-distribution with $n-2$ degrees of freedom. Again the predefined level of significance will be at $\alpha = 5\%$.

1.4.3 Empirical Results

30 HSI Constituents. Results of Q statistics and p-values for the driving noise residuals at lags from order 1 to 16 are shown in table 1.1. At 5% level of significance, autocorrelations of all residuals are not significantly different from zero. It implies that the underlying AR(1) model is adequate as the driving noise residuals behave as a white-noise process. On the other hand, out of 435 cross correlation coefficients, 15 of them, or 3.45% are statistically significant at $\alpha = 5\%$. As the percentage is quite small, the results are satisfactory and we accept the null hypothesis that the observation noise residuals are white. Partial results showing t-test on correlation coefficients the first two stocks with respect to the 30 HSI constituents are shown in table 1.2. Results of the other 28 stocks are omitted due to space constraint.

32 HSCCI Constituents. Results are shown in table 1.3. At 5% level of significance, autocorrelations of all residuals are not significantly different from zero at lag orders 1-16. On the other hand, out of 496 cross correlation coefficients, only 16 of them, or 3.23% are significant at $\alpha = 5\%$. Partial results for the first two stocks are shown in table 1.4. The null hypothesis is accepted based on the results.

Table 1.1. Results showing Q-statistics and the respective p-value of the residuals for 30 HSI constituents.

Lag	Q-Stat Residual 1	p-value	Q-Stat Residual 2	p-value
1	0.0053	0.9418	0.0027	0.9585
2	0.5971	0.7419	0.1644	0.9211
3	2.1562	0.5406	1.9642	0.5799
4	2.6474	0.6185	2.2831	0.6838
5	3.1348	0.6792	2.6673	0.7511
6	3.1836	0.7855	3.1282	0.7926
7	4.9165	0.6701	3.1850	0.8674
8	5.2077	0.7352	3.2273	0.9193
9	5.3895	0.7991	4.4875	0.8765
10	5.5636	0.8505	6.4624	0.7750
11	5.5643	0.9008	7.0889	0.7918
12	6.7370	0.8745	7.4907	0.8235
13	8.0008	0.8435	7.8025	0.8562
14	8.0028	0.8892	7.9810	0.8903
15	8.3492	0.9090	8.0615	0.9213
16	8.3508	0.9377	8.6337	0.9277

Lag	Q-Stat Residual 3	p-value	Q-Stat Residual 4	p-value
1	0.0154	0.9012	0.0001	0.9914
2	0.7600	0.6839	0.0165	0.9918
3	0.7705	0.8565	0.2278	0.9730
4	0.8643	0.9296	0.3707	0.9848
5	1.0434	0.9590	7.4297	0.1906
6	1.0691	0.9829	8.5998	0.1974
7	1.4115	0.9852	8.9039	0.2597
8	1.5148	0.9925	8.9575	0.3459
9	2.1436	0.9890	10.4879	0.3125
10	2.1624	0.9949	10.6861	0.3825
11	2.4104	0.9965	10.9358	0.4487
12	2.4132	0.9985	13.0340	0.3666
13	4.2491	0.9882	15.1241	0.2997
14	4.3633	0.9928	19.8678	0.1344
15	5.9443	0.9807	20.5308	0.1526
16	5.9568	0.9885	20.7737	0.1875

Table 1.2. Partial results of t-test on the observation noise residuals for 30 HSI constituents. Only correlation coefficients the first two stocks with respect to 30 constituents are shown. Results of the other 28 stocks are omitted due to space constraint.

Stock #	corr. Stock	t-stat. #1	p-value	corr. Stock	t-stat. #2	p-value
1	1.0000	–	–	-0.0041	-0.0939	0.9252
2	-0.0041	-0.0939	0.9252	1.0000	–	–
3	-0.0682	-1.5571	0.1201	-0.0215	-0.4888	0.6252
4	-0.0047	-0.1081	0.9140	0.0448	1.0212	0.3076
5	0.0275	0.6268	0.5311	-0.0041	-0.0927	0.9261
6	-0.0381	-0.8677	0.3860	-0.0112	-0.2554	0.7985
7	-0.0699	-1.5958	0.1111	0.0619	1.4133	0.1582
8	-0.0090	-0.2061	0.8368	-0.0115	-0.2609	0.7943
9	-0.0251	-0.5728	0.5670	0.0275	0.6272	0.5308
10	0.0408	0.9294	0.3531	-0.0606	-1.3840	0.1669
11	0.0307	0.7002	0.4841	0.0073	0.1674	0.8671
12	0.0001	0.0032	0.9975	-0.0138	-0.3143	0.7534
13	-0.0257	-0.5863	0.5580	-0.0217	-0.4949	0.6209
14	-0.0834	-1.9058	0.0572	-0.0270	-0.6151	0.5388
15	-0.0978	-2.2382	0.0256	-0.0303	-0.6903	0.4903
16	-0.0516	-1.1776	0.2395	0.0016	0.0373	0.9703
17	0.0821	1.8764	0.0612	0.0571	1.3030	0.1932
18	-0.0439	-1.0007	0.3175	-0.0573	-1.3075	0.1916
19	-0.0341	-0.7784	0.4367	0.0283	0.6456	0.5188
20	-0.0069	-0.1583	0.8743	-0.0484	-1.1036	0.2703
21	-0.0272	-0.6192	0.5360	0.0555	1.2662	0.2060
22	0.0115	0.2630	0.7927	0.0840	1.9205	0.0553
23	0.0567	1.2932	0.1965	-0.0156	-0.3554	0.7224
24	-0.0202	-0.4604	0.6454	0.0645	1.4719	0.1417
25	-0.0557	-1.2702	0.2046	0.1123	2.5758	0.0103
26	-0.0580	-1.3240	0.1861	0.0394	0.8982	0.3695
27	0.0139	0.3156	0.7524	-0.0030	-0.0690	0.9450
28	0.0341	0.7772	0.4374	0.0428	0.9769	0.3291
29	0.0294	0.6694	0.5035	0.0268	0.6115	0.5411
30	-0.0156	-0.3549	0.7228	-0.0394	-0.8982	0.3695

24 HSCEI Constituents. Results are shown in table 1.5. At 5% level of significance, autocorrelations of all driving noise residuals are not significantly different from zero at lags order 1-16. On the other hand, out of 276 cross correlation coefficients, only 11 of them, or 3.99% are significant at $\alpha = 5\%$. Partial results for the first two stocks are shown in table 1.6. The null hypothesis is accepted based on the results.

All 86 Securities. Results are shown in table 1.7. At 5% level of significance, autocorrelations of all residuals are not significantly different from zero at lag orders 1-16. It implies that the underlying AR(1) model is adequate as the driving noise residuals behave as a white-noise process. On the other

Table 1.3. Results showing Q-statistic and p-value of the driving noise residuals for 32 HSCCI constituents.

Lag	Q-Stat Residual 1	p-value	Q-Stat Residual 2	p-value	Q-Stat Residual 3	p-value
1	0.0165	0.8978	0.0397	0.8421	1.2106	0.2712
2	1.6987	0.4277	0.7411	0.6903	1.2741	0.5289
3	1.7028	0.6363	0.7590	0.8592	2.5560	0.4653
4	2.4884	0.6467	3.8699	0.4239	2.5579	0.6343
5	3.1986	0.6694	4.3257	0.5036	3.4214	0.6353
6	4.5150	0.6073	6.7944	0.3403	4.0612	0.6684
7	7.2552	0.4028	6.8038	0.4496	4.2960	0.7451
8	7.5093	0.4828	6.9982	0.5368	6.9552	0.5415
9	13.5159	0.1407	15.2182	0.0852	8.9411	0.4428
10	14.2382	0.1625	18.0936	0.0535	10.8558	0.3689
11	14.4412	0.2096	18.4788	0.0712	11.3892	0.4113
12	14.4568	0.2726	18.4793	0.1020	11.8557	0.4574
13	14.6595	0.3291	21.4853	0.0639	11.9348	0.5330
14	15.2772	0.3595	21.7078	0.0849	14.4735	0.4151
15	15.2816	0.4314	21.7519	0.1146	15.0014	0.4514
16	16.4103	0.4248	22.0868	0.1405	15.0131	0.5237

hand, out of 3655 cross correlation coefficients, only 87 of them, or 2.38% are significant at $\alpha = 5\%$. The null hypothesis is accepted based on the results.

1.5 Synthesis of Economic Factors

Learning via the adaptive TFA algorithm guarantees the recovered Gaussian temporal factors statistically uncorrelated. The constraint of statistical independence is important because it would eliminate the possibility of more than one solution fitting the model. Nonetheless, interpretation of the recovered uncorrelated statistical factors may be difficult as time series of common economic variables are often correlated to some degree. Moreover, as the finance community is more inclined to accept macroeconomic variables as the hidden driving force of stock returns because of both intuition and empirical evidence, attempt to build up a relationship between the recovered statistical factors and some well-known macroeconomic factors is crucial for both the theoretical analysis and practical application of the APT model. We refer to the exploration of the relationship between the economic factors and the statistically uncorrelated factors in this paper the synthesis of APT economic factors.

Admittedly, it is not easy to tell what macroeconomic time series may be treated as the most appropriate APT economic factors. Since the search for the most suitable candidates from a vast number of possible time series is always far from exhaustive, there can be no definite answer to the question

Table 1.4. Partial results of t-test on the observation noise residuals for 32 HSCCI constituents. Only correlation coefficients the first two stocks with respect to 32 constituents are shown. Results of the other 30 stocks are omitted due to space constraint.

Stock #	corr. Stock	t-stat. #1	p-value	corr. Stock	t-stat. #2	p-value
1	1.0000	–	–	-0.0693	-1.5827	0.1141
2	-0.0693	-1.5827	0.1141	1.0000	–	–
3	-0.0085	-0.1944	0.8460	-0.0160	-0.3650	0.7153
4	-0.0096	-0.2192	0.8265	-0.0120	-0.2727	0.7852
5	-0.0296	-0.6756	0.4996	-0.0372	-0.8482	0.3967
6	-0.0418	-0.9541	0.3405	-0.0634	-1.4465	0.1486
7	-0.0038	-0.0874	0.9304	0.0075	0.1717	0.8637
8	0.0162	0.3698	0.7117	-0.0040	-0.0906	0.9278
9	0.0054	0.1221	0.9029	0.0387	0.8829	0.3777
10	-0.0386	-0.8792	0.3797	0.0434	0.9899	0.3227
11	0.0046	0.1050	0.9164	-0.0026	-0.0592	0.9528
12	0.0092	0.2106	0.8333	0.0153	0.3497	0.7267
13	-0.0363	-0.8281	0.4080	0.0410	0.9338	0.3508
14	-0.0045	-0.1018	0.9189	-0.0162	-0.3689	0.7123
15	-0.0358	-0.8162	0.4148	-0.0080	-0.1821	0.8556
16	0.0148	0.3373	0.7360	0.0067	0.1522	0.8791
17	0.0716	1.6357	0.1025	-0.0359	-0.8195	0.4129
18	-0.0350	-0.7985	0.4249	0.0169	0.3845	0.7007
19	-0.0135	-0.3075	0.7586	-0.0260	-0.5922	0.5540
20	0.0261	0.5939	0.5529	0.0315	0.7180	0.4731
21	-0.0045	-0.1018	0.9190	-0.0777	-1.7750	0.0765
22	-0.0565	-1.2881	0.1983	0.0056	0.1272	0.8988
23	0.0151	0.3440	0.7310	0.0359	0.8176	0.4140
24	0.0534	1.2187	0.2235	0.0087	0.1992	0.8422
25	0.0320	0.7285	0.4666	-0.0104	-0.2373	0.8125
26	-0.0264	-0.6016	0.5477	0.0305	0.6943	0.4878
27	0.0222	0.5069	0.6125	-0.0241	-0.5500	0.5825
28	-0.0638	-1.4555	0.1461	0.0279	0.6368	0.5245
29	0.0476	1.0862	0.2779	-0.0202	-0.4612	0.6449
30	0.0103	0.2336	0.8154	0.0489	1.1153	0.2652
31	0.0508	1.1587	0.2471	-0.0296	-0.6735	0.5009
32	-0.0266	-0.6054	0.5452	-0.0188	-0.4283	0.6686

Table 1.5. Results showing Q-statistic and p-value of the driving noise residuals for 24 HSCEI constituents.

Lag	Q-Stat Residual 1	p-value	Q-Stat Residual 2	p-value
1	2.0511	0.1521	0.0639	0.8004
2	4.4732	0.1068	0.0650	0.9680
3	5.9503	0.1141	0.1283	0.9882
4	5.9518	0.2028	0.7236	0.9484
5	5.9692	0.3093	1.7719	0.8797
6	6.9439	0.3261	1.8497	0.9330
7	7.4245	0.3861	4.3465	0.7391
8	7.5826	0.4753	5.0000	0.7576
9	9.4875	0.3936	6.1129	0.7286
10	10.1511	0.4274	6.5875	0.7637
11	10.8101	0.4593	9.9673	0.5333
12	12.4066	0.4136	12.1570	0.4332
13	14.4841	0.3407	14.8576	0.3164
14	15.5100	0.3443	16.1930	0.3018
15	15.9087	0.3882	16.3040	0.3622
16	17.9654	0.3260	20.3909	0.2032

Lag	Q-Stat Residual 3	p-value	Q-Stat Residual 4	p-value
1	0.0048	0.9447	0.0110	0.9166
2	0.4872	0.7838	0.4207	0.8103
3	0.5800	0.9010	0.4266	0.9347
4	8.1902	0.0849	5.4541	0.2438
5	10.2321	0.0689	5.5303	0.3547
6	10.3302	0.1115	8.3598	0.2129
7	10.7253	0.1511	8.5396	0.2875
8	10.7254	0.2178	9.4638	0.3047
9	10.8287	0.2877	10.2432	0.3312
10	10.9531	0.3612	11.7718	0.3007
11	11.2581	0.4219	15.6198	0.1559
12	11.4262	0.4928	17.6307	0.1274
13	12.1782	0.5131	17.6722	0.1704
14	12.6480	0.5544	18.4682	0.1864
15	12.7657	0.6204	19.0036	0.2137
16	14.2880	0.5773	19.3154	0.2527

Table 1.6. Partial results of t-test on the observation noise residuals for 24 HSCEI constituents. Only correlation coefficients the first two stocks with respect to 24 constituents are shown. Results of the other 22 stocks are omitted due to space constraint.

Stock #	corr. Stock	t-stat. #1	p-value	corr. Stock	t-stat. #2	p-value
1	1.0000	–	–	0.0182	0.4150	0.6783
2	0.0182	0.4150	0.6783	1.0000	–	–
3	0.0076	0.1735	0.8623	-0.0108	-0.2463	0.8055
4	0.0094	0.2134	0.8311	-0.0251	-0.5716	0.5678
5	-0.0317	-0.7222	0.4705	0.0245	0.5577	0.5773
6	-0.0111	-0.2519	0.8012	-0.0487	-1.1100	0.2675
7	-0.0173	-0.3933	0.6943	0.0055	0.1260	0.8998
8	0.0089	0.2019	0.8400	-0.0133	-0.3040	0.7613
9	-0.0512	-1.1670	0.2437	0.0609	1.3895	0.1653
10	0.0738	1.6852	0.0925	0.0282	0.6425	0.5209
11	0.0271	0.6172	0.5374	-0.0156	-0.3554	0.7224
12	-0.0090	-0.2056	0.8372	0.0369	0.8415	0.4004
13	-0.0114	-0.2602	0.7948	0.0201	0.4571	0.6478
14	0.0611	1.3954	0.1635	-0.0030	-0.0673	0.9464
15	-0.0022	-0.0504	0.9598	0.0199	0.4525	0.6511
16	-0.0047	-0.1063	0.9154	0.0555	1.2670	0.2057
17	0.0337	0.7671	0.4434	-0.0054	-0.1232	0.9020
18	-0.0112	-0.2553	0.7986	0.0082	0.1873	0.8515
19	0.0146	0.3321	0.7399	-0.0506	-1.1542	0.2490
20	0.0220	0.5002	0.6171	-0.0688	-1.5702	0.1170
21	0.0088	0.2008	0.8409	0.0180	0.4091	0.6826
22	0.0378	0.8624	0.3889	0.0713	1.6281	0.1041
23	0.1204	2.7629	0.0059	0.0233	0.5300	0.5963
24	0.0372	0.8476	0.3971	-0.0140	-0.3193	0.7496

Table 1.7. Results showing Q-statistics and the respective p-value of the residuals for all 86 securities.

Lag	Q-Stat Residual 1	p-value	Q-Stat Residual 2	p-value
1	0.0743	0.7852	0.2992	0.5844
2	1.8990	0.3869	1.2295	0.5408
3	3.7295	0.2922	1.4767	0.6877
4	4.5944	0.3315	1.5594	0.8161
5	5.5482	0.3527	4.8766	0.4311
6	8.0917	0.2315	4.9022	0.5564
7	8.1760	0.3174	8.9754	0.2545
8	8.3028	0.4045	9.0244	0.3403
9	8.6708	0.4682	9.3282	0.4076
10	10.8504	0.3693	11.5515	0.3162
11	13.7944	0.2446	13.9240	0.2373
12	14.7830	0.2536	14.3835	0.2770
13	14.7841	0.3211	18.3171	0.1459
14	16.1057	0.3070	20.0656	0.1282
15	16.1238	0.3739	20.7264	0.1459
16	16.2176	0.4379	22.9507	0.1151

Lag	Q-Stat Residual 3	p-value	Q-Stat Residual 4	p-value	Q-Stat Residual 5	p-value
1	0.0005	0.9818	0.0795	0.7780	3.5885	0.0582
2	0.0666	0.9672	3.9236	0.1406	3.6315	0.1627
3	0.0733	0.9948	3.9439	0.2676	3.9939	0.2621
4	8.9685	0.0619	4.2149	0.3777	4.1009	0.3925
5	9.0195	0.1083	6.8929	0.2288	4.1699	0.5252
6	9.0344	0.1717	7.7789	0.2548	4.1714	0.6535
7	10.6932	0.1526	9.1086	0.2450	7.4628	0.3824
8	11.5813	0.1709	11.4475	0.1777	10.6244	0.2240
9	11.5813	0.2380	11.5153	0.2421	11.5409	0.2405
10	12.0479	0.2819	12.4976	0.2532	11.6855	0.3067
11	12.0507	0.3599	16.5561	0.1218	11.7112	0.3858
12	12.6850	0.3924	17.0175	0.1490	12.3739	0.4162
13	12.9104	0.4548	19.5611	0.1068	12.5800	0.4808
14	12.9815	0.5280	19.7313	0.1389	12.6173	0.5569
15	13.1612	0.5899	20.2641	0.1621	12.6222	0.6314
16	13.2102	0.6573	20.7273	0.1893	13.1958	0.6584

of whether a specific economic time series should be the optimum candidate. Moreover, it is possible that the so-called economic factors such as inflation rate, interest rate and unemployment rate are also being affected by some hidden driving force. This argument is similar to that stock returns can be conceived as being affected by some unknown hidden factors. Broadly speaking, security returns can also be considered as sorts of macroeconomic time series. Thus, if given sufficient macroeconomic time series, it may be possible that TFA can be used to recover those hidden common independent factors for those macroeconomic time series and comparison can be made on their correlations with those statistically independent factors affecting stock returns.

1.5.1 Methodology and Test Statistics

Since we are only able to collect few economic time series, the proposition of factor correlations comparison seems not viable. An alternative to explore the relationship between economic factors and the statistically uncorrelated factors is by means of statistical test on the significance of the coefficients of a regression between an economic time series and the statistically independent factors extracted from stock returns series via TFA. As usual, the time series have been preprocessed so that stationarity is guaranteed. The t-statistic would be used to test for the individual significance while the F-statistic could be used to test for the joint significance of the coefficients. We will examine the results of both tests at levels of significance of 5%. The null hypothesis is that all factor loadings are simultaneously zero. Thus the alternate hypothesis is H_1 : There exist nonzero constants $a_1, a_2, \cdots, a_k$ such that

$$TS_t^{(j)} = a_1^{(j)} y_{1t} + a_2^{(j)} y_{2t} + \cdots + a_k^{(j)} y_{kt} \tag{1.15}$$

where $TS_t^{(j)}$ is the value of the j-th transformed time series at time t.

1.5.2 Empirical Results

We have used four economic time series during the same period from January 1, 1998 to December 31, 1999 for empirical test. They are the 1 month Hong Kong Inter-Bank Middle Rate (Series A), 1 year Hong Kong Inter-Bank Middle Rate (Series B), the Hang Seng Index (Series C) and the Dow Jones Industrial Average (Series D) respectively. We use the five independent factors recovered from all 86 securities for regression. The results showing the p-values of each coefficient are shown in table 1.8. Since at $\alpha = 5\%$ the coefficients are both individually and jointly significant, the null hypothesis is rejected and we can reasonably conclude that there is linear relationship between the time series under test and the statistically uncorrelated factors recovered by TFA.

Table 1.8. Results of t-test and F-test of regression coefficients for real economic time series.

	a_1	p-value a_2	of a_3	t-test a_4	a_5	p value of F-test
A	0.0000	0.0011	0.0000	0.0000	0.0006	0.0000
B	0.0001	0.0001	0.0000	0.0005	0.0000	0.0000
C	0.0003	0.0000	0.0071	0.0000	0.0032	0.0000
D	0.0005	0.0023	0.0001	0.0009	0.0103	0.0000

1.6 Conclusion

We have carried out white noise tests on the residual of the TFA model for model adequacy. The results provide assurance for further statistical analysis using the TFA model. Based on the statistical test results, the null hypothesis is rejected and we accept the alternative hypothesis that each of the four economic time series is linearly related to the statistically uncorrelated factors determined via the TFA model. Therefore APT economic factors can be synthesized from uncorrelated Gaussian temporal factors determined via the TFA model.

References

1.1 Berry, M., Burmeister, E., McElroy, M. (1988): Sorting out risks using known APT factors. Financial Analysts Journal **44**, 29–42
1.2 Chen, N.F., Roll, R., Ross, S. (1986): Economic forces and the stock market. Journal of Business **59**, 383–403
1.3 Estep, T., Hansen, N., Johnson, C. (1983): Sources of value and risk in common stocks. Journal of Portfolio Management **9**, 5–13
1.4 Fama, E., (1970): Efficient capital markets: a review of theory and empirical work. Journal of Finance **25**, 383-417
1.5 La Porta, R., Lakonishok, J., Shliefer, A., Vishny, R. (1997): Good news for value stocks: further evidence on market efficiency. Journal of Finance **52**, 859-874
1.6 Lehmann, B.N., Modest, D.M. (1988): The empirical foundations of the arbitrage pricing theory. Journal of Financial Economics **21**, 213–254
1.7 Roll, R., (1988): R^2. Journal of Finance **43**, 541-566
1.8 Roll, R., Ross, S. (1980): An empirical investigation of the arbitrage pricing theory. Journal of Finance **35**, 1073–1103
1.9 Roll, R., Ross, S. (1984): The arbitrage pricing theory approach to strategic portfolio planning. Financial Analysts Journal **40**, 14–26
1.10 Ross, S. (1976): The arbitrage theory of capital asset pricing. Journal of Economic Theory **13**, 341–360
1.11 Shanken, J. (1982): The arbitrage pricing theory: is it testable. Journal of Finance **37**, 1129–1140

1.12 Xu, L. (1998): Bayesian Ying-Yang learning theory for data dimension reduction and determination. Journal of Computational Intelligence in Finance **6**(5), 6–18
1.13 Xu, L. (2000): Temporal BYY learning for state space approach, hidden markov model and blind source separation. IEEE Transactions on Signal Processing **48**, 2132–2144
1.14 Xu, L. (2001): BYY harmony learning, independent state space and generalized APT financial Analysis. IEEE Transactions on Neural Networks **12**(4), 822–849
1.15 Xu, L. (2001): Best harmony, unified RPCL and automated model selection for unsupervised and supervised learning on Gaussian mixtures, three-layer nets and ME-RBF-SVM models. International Journal of Neural Systems **11**, 43–69