JUSSI TOLVI

Outliers and Predictability in Monthly Stock Market Index Returns*

ABSTRACT

The predictability of monthly stock market index returns is considered, with special emphasis on possible outliers in the data. The data consists of indices from fifteen OECD countries. Three models are used for predicting these series. A random walk model is used as a benchmark, which any other model should be able to improve on. A basic autoregressive (AR) model is then compared with an AR-outlier model, where dummy variables are added for detected outliers. It is assumed that if such outliers are taken into account in the model, the predictions should become better compared to the basic AR model, in which any potential outliers are ignored. Four criteria are used to compare the predictions, a number of which are made both one and two steps ahead. The results indicate that for autocorrelated series (ten out of a total of fifteen series) the outlier model does indeed provide the best predictions.

Keywords: *Forecasting, ARMA model, intervention analysis.*

* An earlier version of this paper was presented at the 20th International Symposium on Forecasting, in Lisbon, June 2000. I also thank an anonymous referee for comments that significantly improved this article.

JUSSI TOLVI, Ph.D. (Econ)
University of Turku, Department of Economics • E-mail: jussi.tolvi@utu.fi

1. INTRODUCTION

The statistical properties of stock market prices have been the subject of a vast literature. A basic assumption is that price movements, or returns, are independently and identically (i.i.d.) distributed. The normal distribution is often used in this research as a benchmark, against which various extensions are considered. Usually in empirical work the i.i.d. and the normality assumptions are rejected, however. More specifically, the observed data has more skewness and kurtosis than normally distributed data, and often significant autocorrelation as well. In addition, the volatility of the return series is usually not constant either, and it is often found that there is considerable dependence in the conditional variance of the data. Similar results are obtained for other financial time series as well. An introduction to these themes can be found in, for example, Campbell, Lo, and MacKinlay (1997).

The predictability of stock market returns, of either individual stocks or indices, is another old research topic. For a recent discussion, see for example Leung, Daouk, and Chen (2000). The observed short term autocorrelation of some return series is a well established fact. Potential explanations for this phenomenon, which is in contrast to the usual assumption of efficient markets, are discussed in, for example, Boudoukh, Richardson, and Whitelaw (1994). Whether this observed autocorrelation can be used for profitable prediction of future returns is a more controversial question. The usual conclusion in the extensive literature on the topic seems to be negative.

There is, however, a somewhat neglected aspect to this question, namely whether the results could be biased because of infrequent outliers in the data. Outliers are here defined as isolated large deviations from the majority of the observations, which can be detected as large residuals in an estimated autoregressive model. This idea is based on the assumption that the returns come from a normal distribution most of the time, but that occasionally one observes very large positive and negative returns. These abnormal returns do not originate from the same distribution as most returns, and are therefore modeled as outliers. A similar, somewhat vague definition of outliers has also been used by Boudoukh et al. (1994), in examining the autocorrelations of stock market index returns. Such outliers are also the most obvious cause for the findings of non-normality in financial data.

Perhaps surprisingly, the effects of outliers on time series predictions have only rarely been considered, although this is sometimes advocated, as in Pack (1990). It is another well known fact that the traditional parameter estimates of autoregressive (AR) models, which are the most frequently used models in this kind of work, are biased in the presence of outliers (see, e.g., Bustos & Yohai, 1986). Therefore taking any potential outliers into account should result in more accurate parameter estimates, and consequently improve also the predictions made with these models.

In this paper the effects of taking potential outliers into account will be examined. The observed return series are predicted first ignoring any outliers, and then taking them into account in the model, and by comparing the resulting predictions. It is often thought that the best comparison between competing models is based on the accuracy of their predictions of the future (out of the estimation sample). Three simple models will therefore be used in this paper in predicting monthly stock market index returns for fifteen OECD countries. The benchmark model is that of a random walk with drift (i.e. no predictability apart from the mean of the series), an AR model, and an AR-outlier model. These models are defined in the next section. The results are then compared using several criteria for the obtained predictions.

This exercise can also be considered from another point of view. Outliers will first be detected from the data, and that knowledge used in prediction. Conversely, it could be argued that this is also a check of whether some observations really are outliers. If the predictions from the outlier model are better, it can be concluded that the observations in question really are not part of the normal data generating process of the series, and their effects should indeed be removed. If, on the other hand, better predictions are obtained by considering no observations as outliers, it could be argued that there are no true outliers in the data. Of course, this is only possible if the underlying model is known. In reality one can only try to approximate the data generating process.

The prediction of outliers themselves, although an interesting question, is something that is left out of consideration in this paper. It would obviously be beneficial if the occurrence of future outliers could somehow be known in advance. However, with the definition of outliers used in this paper, such a task seems impossible. Since outliers are explicitly taken to be separate from the rest of the data, there is no way of predicting them using knowledge only of the past of the same data. Whether some other information could be used for this is outside the scope of this paper.

Other alternatives are also available for modelling data of this kind. An old idea, which has become more popular recently, is to use other distributions than the normal, such as Student's *t*, extreme value and mixture distributions. The downside to such exercises is, however, that often the distributions and related estimation methods are exceedingly difficult to handle. In contrast, the normal distribution is well-known and leads to simple procedures, even when augmented with outliers. Various ARCH and stochastic volatility models are also alternative ways of modelling non-normal data. None of these will, however, be considered in this paper, since dependence in the conditional variance does not affect the point estimates for the conditional mean of the return series, which are needed for the predictions considered here. To simplify matters slightly, the potential effects of the time-varying variance on the outlier detection will be ignored as well.

2. MODELS AND OUTLIERS

The data series in this paper are all returns of stock market indices. Logarithmic differences of the levels of the indices are used, so that for an index series p_t, $t = 0,\dots,T$, where t denotes the time period, the return series is defined as $r_t = \log p_t - \log p_{t-1}$, $t = 1,\dots,T$. The first model to be considered is the simple random walk with a drift in the levels, and therefore the returns are given by

$$r_t = \mu + \varepsilon_t, \tag{1}$$

where μ is the mean of the series, and ε_t is an independently and identically distributed (i.i.d.) error term. This model is used as a benchmark, against which the predictability of the series can be assessed.

The simplest and most commonly used form for introducing dependence is an autoregressive model of order p, denoted AR(p), where

$$r_t = \mu + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + \varepsilon_t. \tag{2}$$

Here, ϕ_j are the autoregressive parameters, and μ and ε_t are as in the random walk model. The third of our models adds outliers to this basic AR model, such that

$$r_t = \mu + \phi_1 r_{t-1} + \dots + \phi_{\mathrm{p}} r_{t-\mathrm{p}} + \varepsilon_t + I[t=k]\omega_k, \tag{3}$$

where $I[t=k]$ is an indicator variable vector, which has a value of one when an outlier occurs at date k, and zero otherwise, and ω_k is the vector of (fixed) magnitudes of the outliers. This model will be called the AR-outlier model.

In the modelling stage, outliers are detected from the data using the basic AR model. In the empirical part of this paper, outliers are defined as observations that have an absolute value larger than three in the basic AR model's standardized residuals. The indicator or dummy variables can then be formed for these observations, based on the detected outliers, and the outlier magnitudes estimated from the data using the AR-outlier model (assuming that the corresponding residual is equal to zero). This is often called intervention modelling in statistical time series analysis (Box & Tiao, 1975). For a more thorough discussion on outliers and their detection, see, for example, Tsay (1988) and Chen and Liu (1993b). Predicting ARMA models in the presence of outliers is examined by Ledolter (1989), Chen and Liu (1993a) and Hotta (1993).

The outlier detection and modelling method outlined above is of course a very simple one. Several more sophisticated methods do exist, such as the iterative ARMA modelling and outlier detection algorithm by Chen and Liu (1993b). Nevertheless, even the simple method used in this paper should provide an indication of whether the predictions are affected by outliers in the data. Boudoukh et al. (1994) have also used a somewhat similar method, but they considered as outliers, and therefore removed from the data, all observations larger (in absolute value) than two times the standard error of the mean of the returns. They found that ignoring such observations did not have much effect on the estimated autocorrelations of their portfolio returns.

3. DATA AND METHODS

The data set used here consists of monthly stock market indices for fifteen OECD countries, and is taken from the OECD database. The span of the data is from January 1970 to August 1999, $T = 356$. The countries and the corresponding indices (and the abbreviations used for them) are Australia, ASE all ordinaries (AU), Austria, VSE WBI (OE), Belgium, BSE all shares (BG), Canada, TSE 300 composite (CN), Denmark, CSE all shares (DK), Finland, HEX all shares (FN), France, SBF 250 (FR), (West) Germany, DAX (BD), Ireland, ISEQ overall index (IR), Japan, TOPIX (JP), The Netherlands, CBS all shares (NL), Spain, MSE general index (ES), Sweden, AFGX (SD), the United Kingdom, FTSE actuaries non-financial (UK) and the United States, Standard & Poor's 500 (US). As mentioned earlier, first differences of the logarithms of the indices are examined in this paper.

Descriptive statistics for the return series are given in Table 1. For a normally distributed variable, skewness is equal to zero, and kurtosis equal to three. The reported normality test is that of Doornik and Hansen (1994), which is a corrected version of the usual Jarque–Bera test, based on the estimated skewness and kurtosis measures. Under the null hypothesis of normality, the test has a χ^2 distribution with two degrees of freedom. As can be seen from the skewness, kurtosis and normality test values, all of the series are clearly non-normally distributed. This is a common finding in financial market data, and is in large part due to a few very large returns, both positive and negative. Such returns will be taken to be outliers in the following, and their effects on the predictability of these returns examined.

Sample autocorrelations, denoted $\hat{r}_j$, for the return series are reported in Table 2 for lags $j = 1,\dots,6$. The standard error of these estimates is $1/\sqrt{T}$, which can be used to assess the statistical significance of the estimates. If the absolute value of an estimated autocorrelation is larger than two times the standard error, or $2/\sqrt{T} = 0.106$, it can be considered significantly different from zero. Of the values in Table 2, the first estimated autocorrelation coefficient is significant for ten of the series, but of the rest only a few isolated ones are significant. Therefore, an AR(1) model is used throughout this paper to model these return series. It is also more or less the standard model used in this situation. Note, however, that the results will be presented separately for the series with statistically significant first lag autocorrelation. This is done to

TABLE 1. Descriptive statistics.

Series	Mean (×100)	Standard deviation	Skewness	Kurtosis	Normality test: statistic	*p*-value
AU	0.560	0.053	-1.499	11.650	105.012	<0.001
OE	0.423	0.049	0.308	8.096	156.775	<0.001
BG	0.536	0.046	-0.910	8.314	95.998	<0.001
CN	0.540	0.050	-0.964	6.627	51.415	<0.001
DK	0.826	0.052	-0.376	6.091	71.676	<0.001
FN	1.129	0.058	-0.317	5.499	53.943	<0.001
FR	0.736	0.059	-0.582	6.349	66.942	<0.001
BD	0.508	0.041	-1.144	8.937	85.765	<0.001
IR	0.943	0.061	-0.630	7.450	100.070	<0.001
JP	0.603	0.042	-0.349	3.841	11.417	0.003
NL	0.508	0.059	-5.351	66.128	499.234	<0.001
ES	0.638	0.056	-0.107	4.907	39.490	<0.001
SD	1.103	0.057	-0.433	5.175	39.517	<0.001
UK	0.836	0.050	0.351	11.809	323.673	<0.001
US	0.748	0.045	-0.708	6.254	55.471	<0.001

TABLE 2. Sample autocorrelations.

Series	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_3$	$\hat{r}_4$	$\hat{r}_5$	$\hat{r}_6$
AU	0.216	−0.011	0.000	−0.050	−0.066	−0.023
OE	0.206	0.112	0.049	0.094	0.055	−0.056
BG	0.235	−0.068	−0.053	−0.009	0.001	−0.011
CN	0.031	−0.064	0.024	−0.053	0.046	0.046
DK	0.035	0.105	0.130	0.123	0.021	0.113
FN	0.229	0.079	0.007	−0.003	0.019	0.070
FR	0.066	−0.080	0.057	0.045	0.044	−0.044
BD	0.294	0.071	0.007	−0.016	−0.041	−0.075
IR	0.211	0.020	0.034	0.025	−0.034	−0.043
JP	0.337	0.069	0.031	0.087	0.005	−0.063
NL	0.055	−0.008	−0.110	−0.036	−0.033	−0.052
ES	0.248	−0.027	−0.010	0.002	0.069	0.061
SD	0.143	−0.020	0.031	−0.017	−0.015	0.028
UK	0.327	−0.028	0.010	0.046	−0.107	−0.139
US	0.011	−0.034	0.005	−0.034	0.085	−0.075

find out whether the misspecification of using an AR(1) model for series with no such autocorrelation affects the results.

In the next section of this paper, predictions, both one and two steps ahead, are computed for samples of 50 observations. This is done such that the first sample consists of observations 1 to 50, the second sample of observations 2 to 51 and so on. This gives in total 306 samples for the one step ahead predictions, and 305 samples for the two step ahead predictions. The random walk model's prediction is simply the estimated (arithmetic) mean ($\hat{\mu}$) of each sample. The basic AR model's one step ahead prediction is $\hat{\mu}+\hat{\phi}r_{t_{50}}$, where $\hat{\mu}$ and $\hat{\phi}$ are the estimated model parameters, and $r_{t_{50}}$ is the last observation of each sample. The two step ahead prediction is $(1+\hat{\phi})\hat{\mu}+\hat{\phi}^2 r_{t_{50}}$. For the AR-outlier model, the basic AR model's standardized residuals are first computed, and for each observation with a residual larger than three in absolute value, an indicator variable is created. The AR-outlier model with these indicator variables is then estimated, and the new parameter estimates used in computing the predictions.

The accuracy of the predictions is compared using four measures. The first three of these are the root mean squared error (RMSE), mean absolute error (MAE) and the median of the absolute errors. These are computed as follows, denoting with r_i^e the prediction for the return at date i. The RMSE is $\sqrt{(\Sigma_i(r_i^e - r_i)^2/n}$, where $i = 1,\dots,n$, and n is the total number of predictions. The MAE is $(\Sigma_i|r_i^e - r_i|)/n$, and median of the absolute errors is the median of $\{|r_i^e - r_i|\}$. Especially the RMSE is sensitive to very large single errors, and therefore the results can be largely determined by only a few observations. In contrast, the median of the absolute errors is robust to large errors in the predictions.

The fourth measure is the percentage of predictions where the correct sign of the future return is predicted. In other words, for each prediction it is determined whether the prediction and the actual return have the same sign, and the percentage of correct predictions is recorded for each model. This is of interest if one only wants to know the future direction of the markets. Note that there are some zero returns in the data. These have been ignored in computing the percentages of correct sign predictions.

The results are also aggregated for all series. The total RMSEs and MAEs are computed for all predictions of all series. For the medians of the absolute errors the aggregate is the mean of the medians for all series. And as already mentioned, the results are also tabulated only for series with statistically significant first lag autocorrelation. All of the computations for this paper were done with the Ox matrix language (Doornik, 1998), using a Pentium PC. The estimation method used for the AR models was exact maximum likelihood.

4. RESULTS

Results for the one step ahead predictions are given in Table 3. The accuracy measures for individual series indicate that no model is always better than the other two. Based on the aggregate results for all series, the results are slightly more promising. The outlier model shares the first position in comparisons based on the first three criteria (RMSE, MAE and Median). Moreover, it is the winner based on the sign criterion.

If only the series with statistically significant autocorrelation are considered, the outlier model emerges as the best model with all four criteria. As can be expected, also the AR model is able to beat the random walk model for these series. It must be kept in mind, however, that the differences between the models are rather small. Note also that both the AR and outlier models improve on the random walk model, based on the sign criterion, irrespective of the series considered (i.e. either all series or autocorrelated series only). In other words, they are able to predict the direction of the next month's price movement more often than the random walk model.

Some nonparametric tests were also computed to see whether these differences in the prediction results are statistically significant, using the obtained accuracy measures for each model as data. The results for the whole data set (all series) do not provide any statistically significant results. Considering the results for the ten autocorrelated series alone, however, there are some statistically significant differences between the models. First of all, pooling the results of all of the four accuracy criteria, the Friedman test for all three models has a p value of 0.007, indicating that the performance of the three models does indeed differ significantly. In pairwise comparisons of the models using the Sign test (not to be confused with the accuracy measure also denoted sign), both of the AR and outlier models appear to be significantly better than the random walk model (with one-sided p values of 0.0125 and 0.0052, respectively), whereas no statistically significant difference is found between the AR and outlier models. Looking next at the results for each of the four accuracy measures separately, the Wilcoxon signed rank test[1] has in comparisons of the outlier versus the random walk model p values of 0.084, 0.12, 0.14, and 0.0054. The test for the sign criterion is therefore clearly significant, the test for the RMSE is significant at the 10% level, and the other two are close to being significant at that level. Of the other pairwise comparisons, the only statistically significant (at the 10% level) difference is between the AR and the random walk models for the sign criterion. In other words, these nonparametric tests also suggest that the outlier model is better than the other two, but not by a clear margin. Keeping in mind the small number of observations available for the tests, this is perhaps not too surprising.

Table 4 tabulates the results for the two step ahead predictions. For this longer prediction horizon, it can be expected that the outlier and AR models lose their predictive power. And indeed, for all series, the random walk model has the best overall predictions according to the RMSE and MAE criteria. The outlier model has a lower value of RMSE than the AR model, but

1 This test takes, in addition to the ranks alone, also the numerical values of the data into account. The test is not appropriate for the pooled data, but results in a more powerful test for the individual accuracy measures.

TABLE 3. Accuracy of the one step ahead predictions: root mean squared error, mean absolute error, median absolute error and percentage of correct sign prediction.

	RMSE			MAE			Median			Sign		
Series	RW	AR	O	RW	AR	O	RW	AR	O	RW	AR	O
AU	5.23	5.37	5.22	3.73	3.81	3.75	2.99	2.93	2.91	56.2	61.0	60.2
OE	5.26	5.31	5.28	3.39	3.37	3.32	1.93	1.79	1.82	54.4	59.6	61.0
BG	4.77	4.67	4.68	3.35	3.24	3.28	2.79	2.61	2.63	50.8	60.4	60.0
CN	5.02	5.11	5.10	3.66	3.74	3.72	3.22	3.12	3.14	54.4	52.0	54.0
DK	5.19	5.24	5.30	3.75	3.82	3.87	2.48	2.61	2.84	59.9	58.6	57.3
FN	5.98	5.89	5.90	4.39	4.28	4.31	3.58	3.32	3.23	56.1	60.7	62.6
FR	6.27	6.38	6.38	4.61	4.67	4.66	3.62	3.67	3.68	58.0	55.4	56.4
BD	4.11	4.06	4.03	2.98	2.93	2.89	2.34	2.22	2.21	57.4	61.0	61.3
IR	6.42	6.45	6.45	4.60	4.64	4.65	3.54	3.79	3.65	56.0	59.4	58.7
JP	4.14	4.04	4.06	3.02	2.94	2.94	2.18	2.04	2.04	63.4	64.1	63.4
NL	6.06	6.35	5.99	3.38	3.48	3.38	2.56	2.50	2.38	56.1	60.7	61.9
ES	5.82	5.73	5.75	4.45	4.43	4.44	3.77	3.86	3.89	55.6	58.0	58.4
SD	5.89	5.98	6.04	4.43	4.56	4.58	3.70	3.78	3.76	58.2	55.8	56.8
UK	5.04	4.93	4.83	3.42	3.42	3.38	2.45	2.55	2.59	61.5	65.8	64.1
US	4.52	4.62	4.68	3.34	3.38	3.39	2.67	2.74	2.72	61.3	58.0	60.0
All series	5.36	5.40	5.36	3.77	3.78	3.77	2.92	2.90	2.90	57.3	59.4	59.8
Auto-correlated series only	5.32	5.30	5.28	3.78	3.76	3.75	2.93	2.89	2.87	57.0	60.6	60.7

Note: The models are a random walk model (RW), an AR(1) model (AR) and an AR-outlier model (O).

a higher value of MAE. Nevertheless, if the median of the prediction errors is used as the criterion, the outlier model is the best model, and the random walk model the worst. Looking at the results for the autocorrelated series alone, the picture is more or less the same. According to the sign criterion, however, the outlier model is the best one, and the random walk model the worst, whether one considers all series or the autocorrelated ones alone. These differences are never statistically significant however, according to the same nonparametric tests that were used for the one step ahead prediction results.

5. DISCUSSION

Three simple models were used to compute one and two step ahead predictions for monthly stock market index returns: a random walk model, an AR(1) model and an AR(1) model with outlier dummies. Overall, the results are somewhat conflicting, but with several accuracy measures the outlier model emerges as the best model for one step ahead predictions, especially if one only considers return series that have statistically significant autocorrelation. Perhaps the

TABLE 4. Accuracy of the two step ahead predictions: root mean squared error, mean absolute error, median absolute error and percentage of correct sign prediction.

	RMSE			MAE			Median			Sign		
Series	RW	AR	O	RW	AR	O	RW	AR	O	RW	AR	O
AU	5.22	5.36	5.29	3.74	3.85	3.79	3.08	3.15	3.17	55.6	57.6	60.4
OE	5.22	5.69	5.52	3.34	3.55	3.54	1.94	1.89	1.87	54.2	54.2	53.8
BG	4.78	4.90	4.94	3.30	3.46	3.50	2.68	2.89	2.90	49.4	50.6	51.9
CN	5.00	5.03	5.05	3.63	3.67	3.69	2.98	3.11	3.09	54.2	54.2	55.0
DK	5.17	5.17	5.21	3.68	3.74	3.82	2.99	2.45	2.69	58.0	59.3	59.3
FN	5.98	6.09	6.15	4.27	4.42	4.46	3.97	3.53	3.55	54.9	56.8	57.3
FR	6.26	6.30	6.29	4.67	4.62	4.62	3.75	3.73	3.56	57.6	56.3	56.6
BD	4.13	4.16	4.16	3.06	3.02	3.01	2.34	2.33	2.37	54.9	56.3	55.3
IR	6.40	6.52	6.54	4.70	4.69	4.71	3.62	3.62	3.65	54.5	56.6	54.5
JP	4.14	4.27	4.28	3.11	3.13	3.12	2.42	2.21	2.20	62.9	61.2	61.6
NL	6.06	6.38	6.09	3.42	3.50	3.43	2.58	2.56	2.41	57.1	55.9	62.6
ES	5.83	6.02	6.01	4.30	4.59	4.61	3.78	3.85	3.84	52.6	53.0	55.8
SD	5.97	5.98	6.01	4.56	4.47	4.50	3.75	3.60	3.58	58.4	60.5	59.5
UK	5.06	5.26	5.26	3.55	3.51	3.48	2.48	2.50	2.42	61.1	63.0	62.0
US	4.58	4.52	4.58	3.47	3.34	3.37	2.71	2.67	2.70	61.5	61.8	60.2
All series	5.37	5.49	5.47	3.79	3.84	3.85	3.00	2.94	2.93	56.5	57.2	57.7
Auto-correlated series only	5.32	5.48	5.47	3.79	3.87	3.87	3.01	2.96	2.96	55.9	57.0	57.2

Note: The models are a random walk model (RW), an AR(1) model (AR) and an AR-outlier model (O).

most significant finding is that the outlier model is, on the whole, able to predict the sign of the next month's return more often than the other two models.

For the outlier model to give better predictions than the random walk model (or the AR model, for that matter), it seems that the series must have at least some degree of autocorrelation. The differences between the models were not great, however. It is therefore unlikely that the minor gain in prediction accuracy could be used for more profitable trading in the stock markets. Nevertheless, the results still indicate that even with such a simple outlier correction at least some improvements in prediction can be obtained. Whether a more sophisticated handling of outliers would improve the results further, remains an open question. Based on the results of this paper, however, this sounds quite plausible. Although not clearly favoured in the results here, the use of outlier modelling should at least not be rejected without further research into its usefulness.

With regard to potential future research, the most obvious extension of this work would be to consider other data sets. Other data frequencies, such as daily and weekly returns, and individual firms' returns could be examined as well. The author has already done some experiments using weekly and daily return data on some indices and some individual stocks. It seems, however, that since in these series there is even less predictability than in the monthly data, predictions tend to be poor, whether or not outliers are taken into account.

In an interesting paper, Frieman and Laibson (1989) present an alternative model for outliers in stock returns. They also discuss the economic implications of outlying stock price movements at some length. Another further topic for research would therefore be the relationships between stock market returns and other economic variables (i.e. fundamentals). For example, some macroeconomic time series are already known to contain outliers (Balke & Fomby, 1994; Tolvi, 2001), and unless outliers in all series are taken into account, the results from regressions between stock returns and other variables may also be misleading. This point has also been raised by Junttila, Larkomaa, and Perttunen (1997), and it could be worthwhile to consider taking outliers into account in using macroeconomic variables in predicting stock returns, and vice versa.

REFERENCES

BALKE, N. S., and FOMBY, T. S. (1994). Large shocks, small shocks, and economic fluctuations: Outliers in macroeconomic time series. *Journal of Applied Econometrics, 9,* 181–200.

BOUDOUKH, J., RICHARDSON, M. P., and WHITELAW, R. F. (1994). A tale of three schools: Insights on autocorrelations of short-horizon stock returns. *Review of Financial Studies, 7,* 539–573.

BOX, G. E. P., and TIAO, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association, 70,* 70–79.

BUSTOS, O. H., and YOHAI, V. J. (1986). Robust estimates for ARIMA models. *Journal of the American Statistical Association, 81,* 155–168.

CAMPBELL, J. Y., LO, A. W., and MACKINLAY, A. C. (1997). *The econometrics of financial markets.* Princeton, New Jersey: Princeton University Press.

CHEN, C,. and LIU, L.-M. (1993a). Forecasting time series with outliers. *Journal of Forecasting, 12,* 13–35.

CHEN, C,. and LIU, L.-M. (1993b). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association, 88,* 284–297.

DOORNIK, J. A. (1998). *Object-oriented matrix programming using Ox 2.0.* London: Timberlake Consultants.

DOORNIK, J. A., and HANSEN, H. (1994). *An omnibus test for univariate and multivariate normality* (An unpublished paper). Nuffield College, Oxford.

FRIEDMAN, B. M., and LAIBSON, D. I. (1989). Economic implications of extraordinary movements in stock prices. *Brookings Papers on Economic Activity, 2/1989,* 137–189.

HOTTA, L. K. (1993). The effect of additive outliers on the estimates from aggregated and disaggregated ARMA models. *International Journal of Forecasting, 9,* 85–93.

JUNTTILA, J., LARKOMAA, P., and PERTTUNEN, J. (1997). The stock market and macroeconomy in Finland in the APT framework. *The Finnish Journal of Business Economics, 4/1997,* 454–473.

LEDOLTER, J. (1989). The effect of additive outliers on the forecasts from ARIMA models. *International Journal of Forecasting, 5,* 231–240.

LEUNG, M. T., DAOUK, H., and CHEN, A.-S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting, 16,* 173–190.

PACK, D. (1990). In defence of ARIMA modeling. *International Journal of Forecasting, 6,* 211–218.

TOLVI, J. (2001). Outliers in eleven Finnish macroeconomic time series. *Finnish Economic Papers, 14,* 14–32.

TSAY, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting, 7*, 1–20.