ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
Lönnrotinkatu 4 B, 00120 Helsinki 12, Finland, tel. 601322

Keskusteluaiheita
Discussion papers

Pekka Ilmakunnas

AGGREGATION VS. DISAGGREGATION IN FORECASTING CONSTRUCTION ACTIVITY*

No 243 08.09.1987

* Presented at the Conference on Disaggregation in Econometric Modelling, Queen's College, Cambridge, April 13th-14th 1987

ISSN 0781-6847

ILMAKUNNAS, Pekka, AGGREGATION VS. DISAGGREGATION IN FORECASTING CONSTRUCTION ACTIVITY. Helsinki : ETLA, Elinkeinoelämän Tutkimuslaitos, The Research Institute of the Finnish Economy, 1987. 20 p. (Keskusteluaiheita, Discussion Papers, ISSN 0781-6847 ; 243).

ABSTRACT: The choice of an appropriate aggregation level is studied in forecasting models for construction activity in Finland. The models use past construction starts to forecast the volume of building construction. In a distributed lag from starts to volume the lag weights can be calculated, based on the way the volume index is calculated. As an alternative, the weights are estimated in a transfer function model. Also ARIMA models are estimated for the volume of construction. The results show that forecasting aggregate volume directly or combining the forecasts for the two main subgroups, housing and non-housing construction, gives good forecast performance. There is in most cases no gain from further disaggregation of non-housing construction to nine subgroups.

KEY WORDS: Distributed lags, ARIMA models, transfer functions, aggregation bias, volume index of construction, construction starts.

1. Introduction

The value of construction at a given period of time originates from construction projects started during the the same period and in past periods. Therefore a forecasting model of construction activity could be based on two submodels: a distributed lag from starts to volume of construction and a forecasting model for started construction. This paper reports on aggregated and disaggregated distributed lag forecasting models of construction activity in Finland. Forecasting starts is not discussed in the paper.

The volume index of construction and information on starts are published in Finland for 10 groups (types) of buildings. Our main interest is, however, in forecasting the volume of total building construction. Also of interest are the largest group, construction of housing (group 0), and the aggregate of the other groups, construction of other buildings (groups 1-9). One has to decide whether to forecast the aggregate index directly or to form disaggregate forecasts for the subgroups which are then aggregated. The traditional argument is that there is loss of information in aggregation. In the present case, there is reason to believe that the subgroup lag patterns contain information that is lost if only an aggregate lag pattern is determined. On the other hand, there is some tradeoff between aggregation and disaggregation since some of the subgroup volume indexes have very wide variations, which seem difficult to forecast. The aggregate index is, in contrast, much smoother and therefore perhaps easier to forecast.

For the subgroups it has been possible to use information on average construction times and costs to calculate the lag weights. This has improved forecasts for the individual groups. For the aggregate series it is possible to calculate a lag structure which is consistently aggregated from the subgroup lags. An alternative approach is to obtain the aggregate forecast from the subgroup forecasts without having to determine the aggregate distributed lag.

In the next section of this paper the calculation of the lag weights and a consistent aggregate lag are explained. In Section 3 the final forecasting models are described and in Section 4 the forecasting performance is studied. Section 5 concludes the paper.

2. Calculation of the lag weights

2.1. Disaggregate lag weights

Information on average constrution times and costs of different types of buildings are used in the calculation of the lag weights. In practice the lengths and costs of projects vary within the groups of buildings. If data on individual projects were available, distributed lags for different projects and the size distribution of the projects could be used for determining the groupwise distributed lags (see Merkies and Bikker (1981) and Trivedi (1985)). In the terminology of Trivedi, the lag distribution is a kernel and the size distribution a compounder. Since the necessary data is not available and the published volume indexes are in any case based on the average times and costs, this aggregation issue is not discussed here. Instead, the paper concentrates on aggregation of the group distributed lags.

Assume that a total of s_{im} cubic meters of construction is started in group i during month m. In the average, these projects last a total of m_i months. The real value (volume) accrued from the projects during the total construction time is

$$Q_{i/m} = \sum_{j=0}^{m_i} p_{ij} s_{im} \tag{1}$$

where p_{ij} is the average real price per cubic meter of buildings in group i in the stage of construction that is under way when $j(j=0,...,m_i)$ months have passed from the start of construction. The data on p_{ij} and m_i can be used for determining the distributed lag from starts to volume in a certain period. van Alphen and Bikker (1976) have used a related method for calculating a distributed lag for the value of housing construction in the Netherlands.

We have information on the costs per cubic meter of different phases of construction and on the average lengths of these phases in months (see Tahvanainen and Lindqvist (1983)). Assume that within a phase, each month has the same share of the cost of the phase. This gives the monthly prices p_{ij}. It is further assumed that no volume accrues at the start of the project so that $p_{i0} = 0$.

The next step is to determine how the volume accrued during a month depends on construction started in past months. During month m the real value of construction in group i is

$$q_{im} = \sum_{j=1}^{m_i} p_{ij} s_{i,m-j} . \tag{2}$$

Hence at the monthly level the distributed lag weights are simply the monthly real construction costs per cubic meter.

The monthly weights still have to be converted to quarterly weights, since the data on the volume and starts are published only quarterly. It is known how many cubic meters of construction are started during a quarter, but not how the starts are distributed within the quarter. If it is assume that starts are evenly distributed, the probability of starts S_{it} (we use upper case letters to denote quarterly figures) to have taken place during any single month of the quarter is $\frac{1}{3}$.

The expected volume accrued in quarter t from projects started during the same quarter is, remembering that $p_{i0} = 0$,

$$[\tfrac{1}{3}(p_{i0} + p_{i1} + p_{i2}) + \tfrac{1}{3}(p_{i0} + p_{i1}) + \tfrac{1}{3}\, p_{i0}]S_{it}$$

$$= \tfrac{1}{3}(2p_{i1} + p_{i2})S_{it} \equiv w_{i0}S_{it}.$$

Similarly, during quarter t projects started at t-1 produce expected volume

$$[\tfrac{1}{3}(p_{i3} + p_{i4} + p_{i5}) + \tfrac{1}{3}(p_{i2} + p_{i3} + p_{i4}) + \tfrac{1}{3}(p_{i1} + p_{i2} + p_{i3})]S_{i,t-1}$$

$$= \tfrac{1}{3}(p_{i1} + 2p_{i2} + 3p_{i3} + 2p_{i4} + p_{i5})S_{i,t-1} \equiv w_{i1}S_{i,t-1}$$

In general form the lag weights are:

$$w_{i0} = \tfrac{1}{3}(2p_{i1} + p_{i2})$$

$$w_{ij} = \tfrac{1}{3}(p_{i,3j-2} + 2p_{i,3j-1} + 3p_{i,3j} + 2p_{i,3j+1} + p_{i,3j+2}),\ j=1,\ldots \quad (3)$$

In the tail of the lag distribution the weights decline because we can assume that $p_{i,3j+k} = 0$ ($k = -2,-1,0,1,2$) when $3j+k > m_i$, i.e. no volume accrues after the end of the average construction time. The weight w_{ij} is zero when $3j-2 > m_i$ or $j > (m_i+2)/3$. The total lag length in quarters, T_i, is the smallest integer value of j for which this inequality holds, minus one. For example, if average construction time is 10 months, volume accrues from construction started in the five quarters t,t-1,...,t-4 and the lag length is four quarters. The sum of the lag weights is $\sum_{j=0}^{T_i} w_{ij} = \sum_{j=1}^{m_i} p_{ij}$, i.e. the total construction cost per cubic meter for a typical project in group i. The volume of construction in quarter t can be expressed as a distributed lag

$$Q_{it} = \sum_{j=0}^{T_i} w_{ij} S_{i,t-j} \tag{4}$$

which can be used for forecasting future values of Q_i when relevant forecasts of S_i are given.

In most subgroups the volume index is actually based on a slightly more detailed disaggregation to different types of buildings and a few size classes for each type. The volume index and amount of starts are published only for the 10 subgroups. Hence in forecasting it is not possible to use more disaggregated distributed lags. After calculating the weights w_{ij} for all size and type classes within each group, these weights in each group were aggregated using the shares of the subclasses in completed buildings (in m^3) in 1982-84.

2.2. Aggregate lag weights

To forecast the total volume of construction activity, it would be desirable to derive a distributed lag from total starts to total volume. In this way only the forecast of total starts would be needed when total volume is forecasted. However, this would ignore variations in the lag patterns between the subgroups. Therefore it is better to start from a consistent aggregate of the subgroup distributed lags. This gives an idea of the probable size of error caused when the intergroup variation is ignored.

Since the group volumes are measured by real value, they can be added to obtain the total volume

$$Q_t = \sum_{i=0}^{9} Q_{it} = \sum_{i=0}^{9} \sum_{j=0}^{T_i} w_{ij} S_{i,t-j} \tag{5}$$

As discussed above, after $j > T_i$, we can set the weights w_{ij} equal to zero. Therefore $T = \max(T_i)$ is used as the common lag length for all groups. This allows the total volume to be written as

$$Q_t = \sum_{j=1}^{T} \sum_{i=0}^{9} w_{ij} S_{i,t-j}$$

$$= 10 \sum_{j=1}^{T} \bar{w}_j \bar{S}_{t-j} + 10 \, \mathrm{Cov}_j(w_{ij}, S_{i,t-j}) \tag{6}$$

where $\bar{w}_j = \frac{1}{10} \sum_i w_{ij}$, $\bar{S}_{t-j} = \frac{1}{10} \sum_i S_{i,t-j}$ and $\mathrm{Cov}_j = \frac{1}{10} \sum_j \sum_i (w_{ij} - \bar{w}_j)(S_{i,t-j} - \bar{S}_{t-j})$.

This kind of decompositions are sometimes used in aggregation theory (see Vartia (1979), van Daal and Merkies (1984)). The first term can be written in the form $\sum_j w_j S_{t-j}$, where $S_{t-j} = \sum S_{i,t-j}$ is total starts in period t-j. This gives a distributed lag from total starts to total volume, where lag weights are averages of the group lag weights. The second term measures the error caused by omission of between group variations. This term would tend to be large when there are always largest amounts of starts in groups where construction costs are highest. There seems to be no reason for this to be always true so that the error from using only the first term may in practice be small.

A second way of decomposing the aggregate volume is

$$Q_t = T \sum_{i=0}^{9} \bar{w}_i \bar{S}_i + T\ Cov_i(w_{ij}, S_{i,t-j}) \tag{7}$$

where $\bar{w}_i = \frac{1}{T} \sum_j w_{ij}$, $\bar{S}_i = \frac{1}{T} \sum_j S_{i,t-j}$ and $Cov_i = \frac{1}{T} \sum\sum_{ij} (w_{ij} - \bar{w}_i)(S_{i,t-j} - \bar{S}_i)$.

The first term can be written as $\sum_i w_i \bar{S}_i$, where $w_i = \sum_j w_{ij}$ is the sum of the lag weights. This is an approximate intertemporal aggregate of the distributed lag, whereas above there was an approximate contemporaneous aggregate. The average of past starts in a group is weighted by the total cost per cubic meter of a typical project in that group. The covariance term now measures the impact of intertemporal variations. In practice using the first term of (7) as an approximation of the lag is not useful, since starts for all 10 groups would still have to be forecasted.

There are several alternative ways of forecasting the aggregate volume. First, the individual group forecasts can be aggregated. If the lag patterns derived above are used as such in forecasting, this yields the same result as when the consistent aggregate lag (5) is used. Second, the lag can be approximated by the first term in (6). Third, one could try to estimate directly a distributed lag from total starts to total volume withoug using the weights w.

2.3. On the adequacy of the calculated weights

In principle it would be possible to use the calculated lag weights directly to forecast the volume of construction. The only necessary adjustment in this case would be to scale the forecasts in each group by average quarterly volume in 1980 in that group. In this way the forecasts are in index form with base year 1980. The group forecasts could then be aggregated using 1980 volume shares as weights. For various reasons this procedure proved out not to be quite satisfactory.

The fitted volume from a distributed lag deviates from the volume calculated by Central Statistical Office for various reasons. First, actual construction times of different projects deviate from the historical group averages. If projects progress consistently faster than on slower than these averages, the actual lag profile differs from the one used here. Differing total construction times have an impact on the tail of the distributed lag; actual lags may be shorter or longer than the average ones. It would be possible to study the distributed lag from starts to completitions to see how the actual lag lengths vary (cf. Borooah (1979)).

Second, the shares of the size classes within a group may vary. For example, shifts towards classes with longer construction times or higher costs would change the actual lag.

Third, the distribution of starts within the quarters may not be uniform. If the actual distribution is concentrated mostly to the beginning (end) of a quarter, the actual lag in quarters is shorter (longer) than the one used here.

Fourth, there may be systematic (seasonal, cyclical or trendlike) changes in the ratio of the volume and the fit from a distributed lag. Seasonality in the ratio may arise from buildings of different sizes being started systematically at different times of the year, or from concentration of starts to certain months e.g. because of holidays. Seasonality was found also by van Alphen and Merkies (1976) in a distributed lag model of the Dutch housing construction. Cyclical variations can be caused by adjustment of construction times to economic conditions. For example, an increase in industrial output may hasten the construction of industrial buildings so that the total construction time is shorter than average and also the lag profile changes. This kind of effects were found in van Alphen and Merkies (1976) and Borooah (1979) in models of housing construction. Trendlike changes in the ratio may result from technical progress which systematically shortens construction times.

Fifth, there are breaks in the available statistics. It was possible to derive time series of starts and volume for 1975-85. However, breaks in the way the Central Statistical Office has calculated the basic data may have changed the relationships of the volume series and the fits from the distributed lags.

The above deficiencies in the calculated lag weights show up as an error term in the models. If the error is completely random, it may not matter in forecasting. However, it is likely that in some cases the error is systematically different from zero and autocorrelated. Therefore it has to be taken into account in the forecasting models.

In some groups the lag weights were changed. The volume index was regressed on the fitted distributed lag and some additional lagged starts. It turned out that in a few cases the lags should be longer than the average construction times had indicated and that more weight should be given to the fail of the lag. This seems to reflect delays in construction, which show up as a peak at the end of the lag profile. The lag weights were corrected in 5 of the 10 groups so that the sum of the weights in each group was kept unchanged.

Seasonality in the lags was taken into account by taking four-quarter differences of the indexes and the fitted lags in the final forecasting models. In the cases where a visual inspection showed that the fit/volume-ratio had changed when there were breaks in the statistics, the estimation period of the forecasting models was shortened. To take into account other changes in the fit/volume-ratio, an error correction term was added into the forecasting models. Finally, also inclusion of business cycle variables was tried, but they did not improve the fits of the models.

3. Forecasting models

Denote the final, corrected, lag weights by w^*_{ij}. The corresponding fit of the distributed lag is $Q^*_{it} = \sum_j w^*_{ij} S_{i,t-j}$. The volume index we want to forecast is $IQ_{it} = 100.Q_{it}/Q_{i80}$, where Q_{i80} is the average quarterly volume of group i in 1980. The estimated models were in the form

$$\Delta^4 IQ_{it} = a\Delta^4 Q^*_{it} + b((Q^*_{i,t-4}/IQ_{i,t-4}) - Z_t) + u_t \qquad (8)$$

where z_t is the "target" value of the ratio Q^*_i/IQ_i. It is defined as the ratio of the 8-quarter moving averages of lagged values of Q^*_i and IQ_i, centered at t-4. The Q^*_i/IQ_i-ratio adjusts in this model to past values of the ratio. The correction term helps to smooth seasonality and errors caused by varying construction times etc. It also accounts for trend-like changes in the Q_i^*/IQ_i-ratio.

This kind of models were estimated with OLS for all 10 subgroups and for the aggregates total building construction and other than housing construction. For the aggregates the model was estimated both using the consistently aggregated distributed lag (5) and the approximation that ignores the covariance term in (6).

As a comparison to the models with a priori determined lag weights, also transfer function models were estimated. In them the lag weights are freely determined. These models have the general form

$$A(L)\Delta^4 IQ_{it} = B(L)\Delta^4 S_{it} + C(L)e_t \qquad (9)$$

where L is a lag operator, A(L), B(L) and C(L) are lag polynomials and e_t is an error term. The orders of the polynomials were identified using cross correlations of the differenced volume indexes and starts, and using the corner method (see Liu and Hanssens (1982)). The final models were estimated with Maximum Likelihood method. Since in most subgroups the estimated lag patterns were not reasonable, only the transfer function models for total building construction, housing construction and construction of other buildings were used. Best forecasting results were obtained when these models did not include contemporaneous starts (i.e. lag 0).

Also ARIMA models were estimated for all 10 subgroups, non-housing construction and total construction. These have the form

$$A(L)\Delta^4 IQ_{it} = C(L)e_t \qquad (10)$$

so that the forecasts are based only on past values of the volume indexes. The models were identified using the autocorrelation, partial autocorrelation and extended autocorrelation functions of the volume indexes (see Liu and Hudak (1983)). The identified models were estimated using Maximum Likelihood. In both transfer function and ARIMA models also the out-of-sample forecasting performance was used as a model choice criterion. This helped in rejecting overparameterized models which had a good fit in the estimation period but did not forecast well.

The transfer function and ARIMA models are used here mainly to compare them with the models with calculated lag weights. However, it is also interesting to study the choice between aggregate and disaggregate forecasting in these models. Theoretically, it can be shown that aggregating disaggregate ARIMA forecasts leads to a smaller mean squared error than forecasting an aggregate time series directly.

It would also be justified to forecast the disaggregate models as a vector ARMA system, where the contemporaneous correlation of the errors is taken into account (see Lütkepohl (1984a)). However, in practice when the parameters and the orders of the lag polynomials are unknown, disaggregation or system estimation do not necessarily improve forecasting performance (see Lütkepohl (1984b)). Also the identification and estimation of a 10-equation vector ARMA model would be a difficult task. Therefore the models have been estimated separately.

In all models the estimation period was 1975.1-1983.4. The actual number of observations in the estimations varied depending on the length of the lags used. The period 1984.1-1985.4 was left for out-of-sample forecasting comparisons. Details on the estimated models are given in Ilmakunnas and Lassila (1987).

4. Forecasting comparisons

No attempts were made to discriminate between aggregation and disaggregation in the estimation period using e.g. tests suggested by Pesaran, Pierse and Kumar (1986). Instead the choice of the forecasting model and the level of aggregation is based on the out of sample forecasting performance of the models.

Three aggregation levels were compared. Total building construction can be forecasted directly, forecasts for housing and non-housing can be aggregated, or forecasts for all 10 subgroups can be aggregated.

Volume shares in 1980 were used as weights when disaggregated forecasts of volume indexes were aggregated. Since in some subgroups the forecasts were quite poor it might actually be possible to change the weights so that groups with better forecasts had more weight. Some alternatives are studied in Ilmakunnas (1986).

Tables 1 and 2 present the forecasting results. Four criteria were used in the comparisons, root mean squared error (RMSE), mean error (ME), mean absolute error (MAE) and mean absolute percentage error (MAPE).

Since the models require as inputs forecasts of starts, two types of forecasts were made. In the first type, it is assumed that the forecast period starts can be perfectly forecasted. This is denoted as forecast horizon 0 in Table 1. In the second type, it is assumed that a "no change" forecast of starts is used. In this case in forecast horizon 1, one period ahead forecasts are made, forecasting $\Delta^4 S_{i,t+1}$ to be zero. In forecast horizon 2, two period ahead forecasts are made, using forecasts $\Delta^4 S_{i,t+1}=0$ and $\Delta^4 S_{i,t+2}=0$. Given these naive forecasts of starts the forecasting performance of the models rapidly deteriorates when the forecast horizon is lenghtened. However, the results for forecast horizon 0 can be used as a comparison to see what can be gained from improving the forecasts of starts. In Ilmakunnas and Lassila (1987) some forecasting models for construction starts are estimated, but in general they do not perform very well.

Since the transfer function models do not include current period starts as explanatory variables, results for forecast horizons 0 and 1 are the same. In forecast horizon 2 it was again assumed that $S_{i,t+1}=0$. The ARIMA models are based on only the past volumes so that the starts need not be forecasted.

According to Table 1, there is not much difference between the volume forecasts when starts are perfectly forecasted or the forecast horizon is short. With a forecast horizon of 2 quarters, aggregation of the forecasts for the two main groups, housing and non-housing construction, and aggregation of all the 10 subgroup forecasts give the best results.

When the total volume of construction is directly forecasted, there is not much difference between using the consistent aggregate lag or an approximate lag structure. This shows that the covariance term in (6) is probably small enough so that it can be left out without worsening the forecasting performance of the model. For forecast horizon 2 the approximate lag is slightly better which probably reflects the effect of having to forecast starts for all the subgroups when the consistent aggregated lag is used.

When forecasts for housing and non-housing construction are aggregated, there is again not much difference between using the consistent lag structure or the approximate lag for non-housing construction. However, if the forecast horizon is 2 periods the approximate lag gives smaller forecast errors. Again this reflects poor forecasts of starts.

Table 2 gives the forecast results for the transfer function and ARIMA models. The aggregate transfer function models gives slightly better results than the distributed lag models for total construction. Hence free estimation of the weights gives better forecasts than the calculated weights. The lag patterns are clearly different, since the transfer function model that was used in obtaining the results includes only starts lagged one and two periods. When the forecast horizon is 2, the transfer function is clearly better than the other models. This may partly be due to the fact that current period starts are not included in the transfer function models.

Combination of the transfer function forecasts for housing and non-housing gives worse results than using the aggregate transfer function forecasts.

Finally, in the ARIMA models the combination of forecasts for housing and non-housing construction is best and the aggregate ARIMA model worst in terms of forecast errors for both forecast horizons. This gives some support to the theoretical results mentioned above, although the optimal level of disaggregation is in this case between the aggregate and completely disaggregate specifications. The forecast performance of the ARIMA models is in general fairly similar to that of the transfer function and distributed lag models.

5. Conclusions

Alternative forecasting models for construction activity in Finland have been formulated and compared at different levels of aggregation. The optimal level of aggregation varies from one model to another. For distributed lag models in which the weights are based on a priori information on the way the volume index of construction is calculated, disaggregate forecasting is preferable only for long forecast horizon. In transfer function models the aggregate model has the best forecast performance. In ARIMA models aggregation of disaggregate forecasts is slightly preferable to aggregate forecasting.

In all, the results show that when the goal is to forecast the aggregate volume of construction, good results may be obtained with fairly simple, even ARIMA, models, and there is not much loss in using aggregate models for forecasting. When the goal is to obtain forecasts for the individual subgroups of construction, the models with calculated lag weights improved forecasting performance in some cases considerably compared to ARIMA forecasts. Detailed results are presented in Ilmakunnas and Lassila (1987).

However, combination of good subgroup forecasts does not necessarily give good forecasts of total construction, since the subgroup forecast errors may not outweigh each other in the average. Also, the aggregate series has relatively less quarter to quarter variation and its changes are therefore more predictable than the disaggregate series. The performance of the combination of disaggregate forecasts may be improved by taking into account the correlations between the error terms of the groups. The good results obtained by combining forecasts of the two main groups, housing and non-housing, may also reflect the fact that when one subgroup (housing) is large compared to the other groups and therefore dominates the movements of the aggregate, there is little gain from further disaggregation of the residual group.

Table 1

Distributed lag models

	Forecast horizon 0	1	2
Forecast of total construction using consistent aggregated lag			
RMSE	2.88	2.80	7.56
ME	.36	.13	-1.49
MAPE	2.38	2.60	6.72
MAE	2.39	2.64	6.73
Forecast of total construction using approximate lag			
RMSE	2.72	2.89	7.13
ME	-.24	-.36	-1.25
MAPE	2.31	2.56	6.14
MAE	2.29	2.53	6.26
Aggregation of forecasts for housing and non-housing; consistent aggregated lag for non-housing			
RMSE	2.61	2.68	5.84
ME	-.26	-.32	-.69
MAPE	2.21	2.38	5.42
MAE	2.31	2.42	5.60
Aggregation of forecasts for housing and non-housing; approximate lag for non-housing			
RMSE	2.96	2.95	3.67
ME	.02	-.01	-1.21
MAPE	2.45	2.45	3.30
MAE	2.55	2.54	3.29
Aggregation of forecasts for 10 subgroups			
RMSE	3.08	3.10	3.82
ME	-.41	-.46	-1.85
MAPE	2.31	2.36	2.86
MAE	2.45	2.47	2.97

Table 2

	Forecast horizon 1	2
Transfer function models		
Forecast of total construction		
RMSE	2.55	2.61
ME	.97	-.21
MAPE	1.80	2.10
MAE	1.92	2.11
Aggregation of forecasts for housing and non-housing		
RMSE	3.08	3.90
ME	.06	-1.52
MAPE	2.60	2.80
MAE	2.68	2.87
ARIMA models		
Forecast of total construction		
RMSE	3.14	3.64
ME	-.93	-1.26
MAPE	2.70	3.10
MAE	2.66	3.02
Aggregation of forecasts for housing and non-housing		
RMSE	2.63	3.18
ME	-1.01	-1.26
MAPE	1.90	2.30
MAE	1.94	2.39
Aggregation of forecasts for 10 subgroups		
RMSE	2.74	3.32
ME	-.98	-1.45
MAPE	2.40	2.90
MAE	2.45	2.90

References:

Alphen, H.J. van and Merkies, H.Q.M.: "Distributed lags in construction: an empirical study", International Economic Review 17, 1976, 411-430.

Borooah, V.K.: "Starts and completitions of private dwellings: four models of distributed lag behaviour", Journal of Economic Studies 6, 1979, 204-215.

Daal, J. van and Merkies, A.H.Q.M.: Aggregation in Economic Research, Dordrecht: Reidel, 1984.

Ilmakunnas, P.: "Aggregation of micro forecasts", in Kähkönen, J. and Ylä-Liedenpohja, J., eds.; A Tribute to Arvi Leponiemi on His 60th Birthday, Acta Academiae Oeconomicae Helsingiensis, Series A:48, 1986, 37-47.

Ilmakunnas, P. and Lassila, J.: Talonrakennustoiminnan ennustemalli, ETLA, 1987 (forthcoming).

Liu, L.-M. and Hanssens, D.M.: "Identification of multiple-input transfer function models", Communications in Statistics A 11, 1982, 297-314.

Liu, L.-M., Hudak, G.B. et al.: The SCA System for Univariate-Multivariate Time Series and General Statistical Analysis, DeKalb, Ill: Scientific Computing Associates, 1983.

Lütkepohl, H.: "Linear transformations of vector ARMA processes", Journal of Econometrics 26, 1984, 283-293. (a)

Lütkepohl, H.: "Forecasting contemporaneously aggregated vector ARMA processes", Journal of Business and Economic Statistics 2, 1984 201-214. (b)

Merkies, H.Q.M. and Bikker, J.A.: "Aggregation of lag patterns with an application in the construction industry", European Economic Review 15, 1981, 385-405.

Pesaran, M.H., Pierse, R.G. and Kumar, M.S.: "On the problem of aggregation in econometrics", Discussion Papers on Structural Analysis of Economic Systems, No. 1, University of Cambridge, 1986.

Tahvanainen, M. and Lindqvist, M.: Volume Index of Building Construction 1980=100, Helsinki: Central Statistical Office of Finland, Studies N:o 88, 1983.

Trivedi, P.K.: "Distributed lags, aggregation and compounding: some econometric implications", Review of Economic Studies 52, 1985, 19-35.

Vartia, Y.: "Kvadraattisten mikroyhtälöiden aggregoinnista", ETLA Discussion Papers, No. 25, 1979.

ELINKEINOELÄMÄN TUTKIMUSLAITOS (ETLA)
The Research Institute of the Finnish Economy
Lönnrotinkatu 4 B, SF-00120 HELSINKI Puh./Tel. (90) 601 322

KESKUSTELUAIHEITA - DISCUSSION PAPERS ISSN 0781-6847

No 215 SEIJA ILMAKUNNAS, The Monopoly Union Model with Endogenous Price Expectations. 15.08.1986. 15 p.

No 216 VESA KANNIAINEN - HANNU HERNESNIEMI, The Cost of Holding Inventories, and the Demand for Labor and Capital under Corporate Taxation: Another Look. 06.10.1986. 24 p.

No 217 TIMO AIRAKSINEN, Pääomaverotuksen teoriaa. 12.11.1986. 63 s.

No 218 VESA KANNIAINEN, Tax Allowances and the Optimal Investment Policy by Firms. 04.12.1986. 45 p.

No 219 JUSSI RAUMOLIN, The Role of Education in the Development of the Mining Sector in Finland. 04.12.1986. 83 p.

No 220 MARKKU RAHIALA - TIMO TERÄSVIRTA - VESA KANNIAINEN, Factors Affecting Firms' Employment Plans in Finnish Manufacturing Industries. 15.12.1986. 30 p.

No 221 TIMO TERÄSVIRTA, Incomplete Ellipsoidal Restrictions in Linear Models. 16.12.1986. 9 p.

No 222 OSMO FORSSELL, Panos-tuotos-laskelmat Suomen Neuvostoliiton-viennistä. 22.12.1986. 119 s.

No 223 OLLI-TAPIO MATTILA, Suomen Neuvostoliiton-kaupan kehitys, kuvioliite. 22.12.1986. 94 s.

No 224 PEKKA ILMAKUNNAS, Survey Expectations vs. Rational Expectations in the Estimation of a Dynamic Model: Demand for Labor in Finnish Manufacturing. 30.12.1986. 22 p.

No 225 PEKKA SPOLANDER, Kapitalmarknader och ägarförhållanden i Finlands näringsliv. 31.12.1986. 42 s.

No 226 JUHA KINNUNEN, Comparison of the Arima-Model Forecasts of Some Finnish Macroeconomic Variables with Econometric Macromodel Forecasts. 31.12.1986. 33 p.

No 227 ERKKI KOSKELA, Personal Savings and Capital Income Taxation: A Differential Incidence Analysis. 12.01.1987. 16 p.

No 228 MORTEN JONASSEN - PAAVO SUNI, Real Exchange Rates as Indicators of Purchasing Power Parity. 20.02.1987. 30 p.

No 229 JUHANI RAATIKAINEN, Variability of Exchange Rates under Rational Expectations. 21.02.1987. 25 p.

No 230 TIMO AIRAKSINEN, Talletusten verollistamisen vaikutus pankkien käyttäytymiseen ja kannattavuuteen. 31.03.1987. 21 s.

No 231 JUHA AHTOLA, Error Correction Mechanism: An Economic Interpretation. 01.04.1987. 10 p.

No 232 HANNU TÖRMÄ, Katsaus eräisiin pohjoismaisiin panossubstituutiotutkimuksiin. 01.04.1987. 49 s.

No 233 HANNU TÖRMÄ, Pääoman, työn, energian ja raaka-aineiden substituutio Suomen, Ruotsin ja Norjan tehdasteollisuudessa. 01.04.1987. 35 s.

No 234 DAVID BENDOR, Finnish Price Competitiveness - A Sectoral Review". 04.06.1987. 70 p.

No 235 VESA KANNIAINEN, An Alternative Corporation Tax: Implications for Efficiency of Investment and Valuations of Shares. 03.06.1987. 17 p.

No 236 PEKKA NYKÄNEN, Tehdasteollisuuden ja sen toimialojen kansainvälinen kilpailukyky. 10.06.1987. 75 s.

No 237 JEAN-PIERRE SICARD - VALDEMAR DOS REIS MEIXEDO, "L'Economie Européenne a l'Horizon 1992. 18.06.1987. 74 p.

No 238 PASI AHDE, Measurement of Capacity Utilization in Manufacturing Industry. 18.06.1987. 22 p.

No 239 PEKKA ILMAKUNNAS, On the Profitability of Using Forecasts. 29.07.1987. 9 p.

No 240 ERKKI KOSKELA, Changes in Tax Progression and Labour Supply under Wage Rate Uncertainty. 06.08.1987. 20 p.

No 241 TIMO TERÄSVIRTA, Superiority Comparisons between Mixed Regression Estimators. 14.08.1987. 11 p.

No 242 SYNNÖVE VUORI, Tiedonhankinnan ja -välityksen kehittäminen Elinkeinoelämän Tutkimuslaitoksessa. 17.08.1987. 54 s.

No 243 PEKKA ILMAKUNNAS, Aggregation vs. Disaggregation in Forecasting Construction Activity. 08.09.1987. 20 p.

Elinkeinoelämän Tutkimuslaitoksen julkaisemat "Keskusteluaiheet" ovat raportteja alustavista tutkimustuloksista ja väliraportteja tekeillä olevista tutkimuksista. Tässä sarjassa julkaistuja monisteita on rajoitetusti saatavissa ETLAn kirjastosta tai ao. tutkijalta.

Papers in this series are reports on preliminary research results and on studies in progress; they can be obtained, on request, by the author's permission.

0033A/08.09.1987