OUTLIER ROBUST ANALYSIS OF MARKET SHARE AND DISTRIBUTION RELATIONS FOR WEEKLY SCANNING DATA

By

Philip Hans FRANSES,[1] Teun KLOEK
Econometric Institute, Erasmus University Rotterdam
P.O.Box 1738, NL-3000 DR Rotterdam, The Netherlands

and

André LUCAS
Department of Financial Sector Management, Free University Amsterdam
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

First Draft: July 29, 1996
Version: July 1996

In this paper we consider empirical econometric models for nine brands of a fast-moving nondurable consumer product using weekly observed scanning data on market share and distribution conditional on advertising, price, and promotion activities. Since the data show nonstationary characteristics, we rely on cointegration techniques to estimate long-run and short-run parameters. Additionally, as there are many outlying observations in our weekly scanning data, we apply robust cointegration methods. We find different results across robust and non-robust methods for the long-run relations between market share and distribution and for the short-run response to disequilibrium situations.

Keywords: market share, distribution, scanning data, cointegration LM-test, outlier robust method.

1. INTRODUCTION

This paper addresses the empirical issue of analyzing weekly observed scanning data at the brand level for such variables as market share and distribution, conditional on prices and various promotional activities. One of the key properties of scanning data, which are observed with a high frequency, is that the data are often contaminated by many irregular observations. Such observations can appear in clusters or may be confined to single data points. For example, for a variable as market share, one cause for the possible presence of outliers in the observation period is that new brands are introduced and that

[1]The first author thanks the Royal Netherlands Academy of Arts and Sciences for its financial support. We thank Willem Verbeke for making the data available to us.

others are withdrawn from the market. Additionally, promotional activities may temporarily boost market shares, but are also well known to result in possible post-promotional dips. For a variable as relative distribution, which is usually defined as some percentage of stores carrying the brand, one may expect measurement errors or sudden out-of-stock situations which may not be easy to explain using an econometric model. As a final example, the price level can show large spikes because of one's own promotional activities or because competitors undertake such activities exactly in the same week or in the week just right after one's own promotions.

It is well known that irregular observations can have large effects on empirical econometric models that describe economic time series variables. These effects range from spuriously significant to spuriously insignificant results, e.g., see Franses and Haldrup (1994), Franses and Lucas (1996), and Lucas (1996a) for recent surveys of these effects in the context of nonstationary time series variables. Such results can concern the parameters that reflect the correlation between the economic variables of interest, but may also concern the time series properties of the variables themselves. Examples of the latter properties are that time series may seem nonstationary (in the sense that the direction of their trend is stochastic and not deterministic), see, e.g., Perron and Vogelsang (1992), or in fact may seem stationary, see, e.g., Lucas (1996a) and Franses and Haldrup (1994), only because of some neglected irregular data points.

For empirical marketing research it is important that one can rely on the availability of guidelines to construct empirical econometric models. See for example a recent issue of Marketing Science (Bass and Wind, 1995), which is dedicated entirely to surveying empirical generalizations in marketing. To enhance the usefulness of such generalizations, say, with respect to the relation between market share and distribution, it is in turn a requirement that they are reasonably robust across brands and sample periods, and hence also robust to irregular observations. Given the possible nature of scanning data as described above, it seems relevant to estimate market share and distribution relationships using procedures that can deal with contaminated data. In this paper we focus on constructing outlier robust econometric models for weekly observed market share and distribution for nine brands of a fast-moving nondurable consumer good.

The outline of our paper is as follows. In Section 2 we present our data and show that these typical scanning data appear to contain many irregular data points. Furthermore, several time series display trending patterns. Since it is uncertain whether these are stochastic or deterministic, we decide to use outlier robust cointegration methods to analyze the data. In the second part of Section 2 we outline the structure of our econometric model, which allows for the possible presence of so-called push and pull effects. In Section 3 we discuss some of the technical details of our outlier robust estimation approach. The material in this section draws upon Lucas (1996b), and for more details we refer the reader to this study. In Section 4 we apply our method to the data for nine brands. We find that the number of long-run relationships can be quite different across robust and non-robust analyses. This emphasizes the hypothesized effects of irregular observations. In the second part of this section we give some details on the long-run relations between market share and distribution and on the short-run response to disequilibrium situations for two selected brands. We find that key parameters can differ across robust and non-robust approaches, indicating that outliers can blur possibly generalizable results on fundamental relations in marketing. In the last section we conclude.

2. DATA AND ECONOMETRIC FRAMEWORK

In this section we start with a discussion of the data and we highlight some of their properties. In Subsection 2.2 we sketch the structure of our econometric time series model to be used in the application.

2.1. *Data*

In this paper we consider the following variables, which are observed weekly:

M_t: Market share;

D_t: Weighted distribution. The distribution of a brand is weighted by the size of a store. The larger stores have a larger weight than the smaller stores. D_t measures the number of stores in which the brand is available;

P_t: The price, relative to all other brands in the same market;

A_t: Share of voice (i.e., the share of the total advertising budget in the market), interpolated from monthly data;

PR_t: Promotion of a certain type, with a value of 1 for a week with promotion, zero otherwise. We distinguish four types of promotion; PR_{1t} is price off and display, PR_{2t} is price off only, PR_{3t} is feature and display, and PR_{4t} is a feature only.

We observe these variables for nine brands, i.e., brands 1, 2, and 3 from companies A, B, and C. These brands are competitors in the same market for a fast-moving consumer good.

Before we analyze these scanning data, we apply some transformations. The market share, distribution, and price observations are transformed by taking natural logarithms, i.e., $m_t = \ln(M_t)$, $d_t = \ln(D_t)$, and $p_t = \ln(P_t)$. A characteristic of the advertising data A_t is that they often take zero values, i.e., for many weeks there is no advertising for one or more of the brands. To construct useful variables, we first calculate

$$a_t = \ln(100(A_t + 0.01)). \tag{1}$$

Then, given the fact that the advertising impact carries over to ubsequent time periods, we generate cumulative advertising data like

$$ca_t = 0.95ca_{t-1} + 0.05a_t, \tag{2}$$

where ca_0 is set equal to a_0. In our empirical models below we include the variable ca_t after the first differencing operator has been applied, i.e., we include $\Delta_1 ca_t = ca_t - ca_{t-1}$. Notice that the transformation of a_t compares with the inclusion of a certain number of lagged untransformed advertising variables in our final models. Hence, with our transformation we aim to reduce the number of regressors. Finally, the promotion dummy variables will also be included in our models after the first differencing operator has been applied.

In this paper we analyze weekly observations for nine brands of the same fast-moving consumer product. For brands A1, B1, B2, C1 and C2, the sample contains 132 weekly observations for the period of week 25 in 1989 through week 52 in 1991. During that period, the brands A3, B3, and C3 were newly introduced. For each of these three novel brands we have data until week 52 in 1991, i.e., the end of the sample. For brands A3, B3, and C3 we have 107, 105, and 86 observations, respectively. Finally, for brand A2 we have 113 data points beginning in week 25 in 1989, since it was withdrawn from the market in week 33 of 1991.

Graphs of the data for m_t, d_t, and p_t for each of the nine brands are displayed in Figures 1 through 9. The graphs for $\Delta_1 PR_{it}$ are not given since these are dummy variables with peaks in only a few weeks. The $\Delta_1 ca_t$ variables display similar features and hence are also not shown. A casual inspection of the graphs for market share, distribution, and prices suggests that for many time series the mean is not constant over time. Market share can increase rapidly, see, e.g., B3, or decrease slowly, see, e.g., A1, B1, and C1. Such slow decreases in market share may be due to the introduction of the novel products of type 3. Distribution may show a cyclical pattern as for C2, although it may also drop exponentially after a certain period as, e.g, for brand A2. The relative price p_t can display a downward trending pattern as for A3, and it may be relatively constant for other brands. All in all, some of the market share, distribution, and price data display nonstationary behavior, and in any subsequent analysis of these time series variables one should take this phenomenon into account.

The second observation to be made from the Figures 1 through 9 is that there seem to be several data points which display irregular behavior. These outlying observations can be caused by changes in promotion activities that temporarily change prices, by the introduction of brands on which we do not have data, by holiday or institutional effects (as, e.g., observation 105 for the distribution data) or by measurement errors. Given the relatively large number of such outlying observations in these scanning data, we feel that robust estimation techniques are useful in our empirical analysis.

2.2. *Remarks on model structure*

In a market for fast-moving consumer products, one may expect the presence of so-called push and pull effects with respect to the marketing efforts to increase sales and possibly market share. The pull effects correspond to the activities of a producer that are directly focused on the consumer, such as advertising and price cuts. The push effects concern the marketing efforts directed towards the retailer, such as promotions of several types. See Farris et al. (1989) for a recent theoretical model for push and pull effects. In this paper, where we focus on the construction of empirical econometric models for market share and distribution, the presence of push and pull effects amounts to a bivariate causality structure between distribution and market share and to the inclusion of marketing activities as advertising and promotion in both models. Hence, for a proper analysis of market share and distribution, we have to take this possible bivariate causality structure into account.

Roughly speaking, we consider the structure

$$m_t = f(d_t, p_t, ca_t, PR_t, m_{t-1}), \tag{3}$$

where f denotes a linear function in this paper. Of course, model (3) can be extended in various ways, e.g, by including additional dynamics, but we do not pursue such extensions in this section for notational convenience. The empirical models in Section 4, however, will include more dynamics. The relevance of p_t, ca_t, and PR_t may be interpreted as pull effects, while m_{t-1} may reflect the brand loyalty of consumers.

Push effects can be observed in a model for distribution, i.e.,

$$d_t = g(d_{t-1}, p_t, ca_t, PR_t, m_t), \tag{4}$$

where g is some linear function. Since distribution is measured as the relative availability of a certain brand, the inclusion of variables as price, advertising, and promotion seems relevant. The d_{t-1} variable in (4) reflects the brand loyalty of the retailer.

Given the apparent trends in the data for m_t, d_t, and p_t, we confine our cointegration analysis to these three variables. Since our model framework should allow for the impact of promotional activities, we include the first differences of the relevant variables, ca_t and PR_{it}, as exogenous variables in our cointegration time series model. This ensures that these variables only have an impact on the level of market shares, distribution levels, and prices, and not on their growth rates. Our focus in this paper is on the (possible) long-run relation between m_t, d_t, and p_t, as well as on the short-run error correction parameters in case there is cointegration. To save space, we do not explicitly consider the effects of short-run variables such as $\Delta_1 ca_t$ and $\Delta_1 PR_{it}$. To obtain proper statistical models, however, we do include all stationary variables in our cointegration model such that our results for long-run parameters are conditional on these short-run variables.

3. OUTLIER ROBUST COINTEGRATION ANALYSIS

The cointegration testing procedure employed in this paper extends the testing procedure put forward in Lucas (1996b). As is well-known in the statistical literature, estimators and testing procedures derived under the assumption of normally distributed errors are very sensitive to outliers, fat-tailedness, and other departures from the normality assumption. Robust estimators and tests are much less sensitive in this respect, while they can be designed such that they are nearly optimal if the normality assumption happens to be true. Therefore, robust procedures seem suitable candidates for testing for cointegration using the scanning data described in Section 2.

3.1. *Preliminaries*

The testing procedure is as follows. Let $y_t \in \mathbb{R}^k$ be a vector of observed variables. In our case we have $k = 3$ and $y_t = (m_t, d_t, p_t)^\top$, where $^\top$ denotes transposition. We assume that y_t follows the vector autoregressive (VAR) model with exogenous variables

$$\Delta_1 y_t = AB^\top y_{t-1} + \mu + \gamma t + \sum_{i=1}^{p-1} \Phi_i \Delta_1 y_{t-i} + \Psi x_t + \varepsilon_t, \tag{5}$$

where the vectors ε_t are independently and identically distributed with mean zero and finite variance-covariance matrix Ω_{11}. The parameter matrices A and B are $k \times r$ matrices of full column rank $r \leq k$. The $k \times k$ matrices Φ_i give the effect of lagged differences of the endogenous variables. μ and γ are the parameters of the deterministic components. Following Johansen (1991, 1994) and Lucas (1996a), $\mu \neq 0$ may cause a linear trend in y_t, while $\gamma \neq 0$ may cause a quadratic drift in y_t. We exclude the possibility of quadratic drift in y_t in the present paper by assuming that $\gamma = A\gamma_0$ for some $\gamma_0 \in \mathbb{R}^r$. Although we do not exploit this information in the estimation stage for simplicity, the assumption of no quadratic trending has major effects on the statistical distribution theory of the cointegration test, see, e.g., Johansen (1994) and Lucas (1996a). Finally, x_t with associated parameter matrix Ψ contains auxiliary regressors, e.g., the promotion and advertising variables $\Delta_1 PR_{it}$ and $\Delta_1 ca_t$.

The cointegration testing procedure of Johansen (1991) now proceeds by assuming that the ε_t are normally distributed. Given this assumption, one can derive a likelihood ratio test for the hypothesis that there are at most r cointegration relations. Other likelihood-based testing procedures like the Wald and the Lagrange Multiplier (LM) test, can also be used, see Kleibergen and van Dijk (1994). As argued before, however, tests based on the assumption of normality are generally nonrobust to aberrant observations. In order to overcome this sensitivity of the normal-based cointegration tests, Lucas (1996a,b, 1997) proposes cointegration tests based on the pseudo maximum likelihood estimation principle, compare Gouriéroux et al. (1984) and White (1982) for the stationary case. The intuition behind these tests is as follows. One postulates a pseudo likelihood by assuming a certain distribution for the innovations ε_t. One then maximizes this pseudo likelihood over the set of unknown parameters. As the pseudo likelihood need no longer coincide with the true likelihood, care must be taken when computing the usual test statistics, because the information matrix equality no longer holds. Note that practitioners already (implicitly) employ the pseudo maximum likelihood principle by using the procedure of Johansen (1991) in settings where the normality assumption for ε_t does not hold, e.g., for financial data. The pseudo likelihood used in that case is the Gaussian one.

The pseudo likelihood-based test we use in this paper is the LM-test. The limiting distribution of this test contains less nuisance parameters than the likelihood ratio (LR) test and, moreover, is less sensitive than the limiting distribution of the Wald test under simple transformations of the variables. For both findings, see Lucas (1996a). Note, however, that even the limiting distribution of the LM-test based on a non-Gaussian pseudo likelihood contains nuisance parameters. These nuisance parameters are the canonical correlations between certain linear combinations of the innovations ε_t and the pseudo score. We use the procedure of Lucas (1996b) to obtain consistent estimates of the nuisance parameters, which can be used in a subsequent inference procedure. Lucas (1996b) shows by means of simulations that this procedure has satisfactory performance in finite samples.

Following Franses and Lucas (1996), we consider a Student t likelihood with 5 degrees of freedom:

$$f(\varepsilon_t) = \frac{\Gamma((5+k)/2)|\Omega_{11}|^{-1/2}}{\Gamma(2\frac{1}{2})\sqrt{3\pi}} \left(1 + \frac{\varepsilon_t^\top \Omega_{11}^{-1} \varepsilon_t}{3}\right)^{-(5+k)/2}. \tag{6}$$

Lucas (1996a) shows that this pseudo likelihood has good properties for both fat-tailed and normal innovations, while Franses and Lucas (1996) show that it has good properties in situations with and without outliers. Moreover, Hoek et al. (1995), using theoretical concepts from the robustness literature, show for the univariate case that the Student t pseudo maximum likelihood estimator with given degrees of freedom parameter is robust. Their proofs can in principle be generalized directly to the present multivariate setting.

3.2. *Our robust cointegration procedure*

In the present paper we suggest a slight and straightforward generalization of the methodology used in the aforementioned papers. The class of pseudo maximum likelihood estimators is a subset of the class of maximum likelihood type (M) estimators, see Hampel et al. (1986) and Huber (1981). These estimators explicitly take account of outliers in ε_t, so-called innovation outliers. They do not, however, explicitly correct for gross-errors, e.g., transcription errors, in y_t itself, so-called additive outliers. In the univariate case, Martin (1979, 1981) therefore suggests to replace the class of M estimators for time series models by the class of generalized M (GM) estimators if one cannot exclude the possibility of additive outliers. Although the M estimator based on the Student t distribution is already robust to additive outliers, see Hoek et al. (1995) and Franses and Lucas (1996), we follow Martin's suggestions to further reduce the influence of outliers on cointegration analysis by employing GM estimators. As the scanning data introduced in Section 2 reveal some highly aberrant patterns, we hope that the use of GM estimators will enhance the interpretability of our empirical findings.

GM estimators assign less weight to observations with either unusually large values of ε_t or unusual values of the regressors. They do so by appropriately centering and scaling the regressors using location and covariance type measures. For models like (5) this causes two types of problems. First, the promotion dummies PR_{it} and the cumulative advertising variable ca_t relate to activities that are given to the econometrician and are, therefore, unlikely to contain observations that are outlying in the usual meaning of the word. Also the constant and the trend do not contain outliers, as these regressors are generated artificially. We therefore exclude x_t, the constant, and the trend from the calculation of the GM weights for the regressors. A second problem with computing the GM weights is the nonstationarity of the regressor y_{t-1} on the right-hand side of (5). As y_{t-1} is nonstationary, we cannot consistently estimate a (time-invariant) covariance type scaling matrix for this regressor. We could solve this problem by calculating the weights based on $B^\top y_{t-1}$, instead. This, however, would make the weights dependent on the unknown parameters B, which would require a simultaneous determination of the GM weights and the unknown parameters. In this paper, we propose a simpler solution. We base the GM weights only on the regressors $\Delta_1 y_{t-i}$ for $i = 1, \ldots, p-1$. This approach has several advantages. First, the GM weights can be determined beforehand instead of simultaneously. Second, the weights do not vary with the postulated cointegration rank r, but only with the order of the VAR model p. The fact that our GM weights are independent of the cointegration relations $B^\top y_{t-1}$ does not affect the robustness of our procedure. As mentioned earlier, even a procedure without GM weights would be robust. Our present use of GM weights only serves to further reduce the impact of aberrant observations on the empirical outcomes. This seems a good strategy for the scanning data of Section 2, some of which show a rather erratic behavior during certain periods.

3.3. *More details*

We now come to a more detailed expostion of the proposed testing procedure. We only sketch the general idea here. For a more precise exposition of the test and the inference procedure along with some limiting distributional results, we refer to Lucas (1996b). We first discuss the construction of the GM weights. Next, we introduce the generalized pseudo likelihood that is to be maximized. Then we give an explicit parameterization of the unit root parameters. This parameterization is used to derive the LM test. Finally, we present some comments on the costs of using outlier robust cointegration analysis if no outliers are present.

First, we discuss the GM weights for the regressors. Following Hampel et al. (1986), we use M estimators of multivariate location and scatter of the vectors $Z_t^\top = (\Delta_1 y_{t-1}^\top, \ldots, \Delta_1 y_{t-p+1}^\top)$. Note that if $p = 1$, the GM weights are absent. The location measure m_Z and covariance measure V_Z are computed using a Huber weighting scheme. Let m_Z^0 and V_Z^0 denote some preliminary estimates of m_Z and V_Z, e.g., the mean and the sample covariance matrix. Next, define the iteration scheme

$$m_Z^{n+1} = \sum_{t=1}^{T} w_{xt}^n Z_t \Big/ \sum_{t=1}^{T} w_{xt}^n, \tag{7}$$

$$V_Z^{n+1} = \sum_{t=1}^{T} w_{xt}^n (Z_t - m_Z^{n+1})(Z_t - m_Z^{n+1})^\top \Bigg/ \left(\sum_{t=1}^{T} w_{xt}^n - 1 \right), \tag{8}$$

where T denotes the sample size, and n denotes the iteration number. The variables w_{xt}^n denote the weights of the tth observation Z_t in the nth iteration. These weights are given by

$$w_{xt}^n = \min\left(1, c_Z((Z_t - m_Z^n)^\top V_Z^n)^{-1}(Z_t - m_Z^n))^{-1}\right), \tag{9}$$

with c_Z the 99th percentile of the χ^2 distribution with $\dim(Z_t)$ degrees of freedom. Other values of c_Z are of course also possible and will result in estimators with different robustness properties. The higher the value of c_Z, the less robust the resulting estimator will be. Note that for $c_Z \to \infty$ we obtain the (highly nonrobust) sample mean and sample covariance matrix, while we also obtain $w_{xt} \equiv 1$ for all values of t. It is easy to see that the weights w_{xt} decline if Z_t becomes exceptionally large compared to the remaining Z_t vectors. (7) and (8) are iterated until m_Z^n and V_Z^n converge. As starting values, we use the coordinate-wise median for m_Z^0, and

$$\sum_{t=1}^{T}(Z_t - m_Z^0)(Z_t - m_Z^0)^\top/(T-1)$$

for V_Z^0. The final GM weights for the regressors are given by $w_{xt} \equiv w_{xt}^\infty$. Note that these weights can be calculated separately from the remaining calculations.

Combining the GM weights for the regressors with the postulated Student $t(5)$ pseudo density from (6), we obtain the generalized pseudo log likelihood

$$\ell(\theta) = c_1 - \sum_{t=1}^{T} w_{xt}\left[\frac{1}{2}\ln|\Omega_{11}| + \frac{(5+k)}{2}\left(1 + \frac{\varepsilon_t^\top \Omega_{11}^{-1}\varepsilon_t}{3}\right)\right], \tag{10}$$

where θ is the vector of unknown parameters, and where c_1 is a constant that does not depend on θ. ε_t in (10) has to be interpreted as a function of the observed values y_t and x_t and the unknown parameters θ. (10) has to be maximized with respect to θ.

In order to derive an LM-test, we need the parameters that are restricted under the null hypothesis. For this purpose, we use the parameterization suggested in Kleibergen and van Dijk (1994), namely $A^\top = (\alpha_{11}^\top, \alpha_{21}^\top)$ and $B^\top = (I_r, \beta^\top)$, with I_r denoting the unit matrix of order r, α_{11} an $r \times r$ matrix, α_{21} a $(k-r)\times r$ matrix, and β a $(k-r)\times r$ matrix. Furthermore, we replace the matrix $AB^\top$ of rank r on the right-hand side of (5) by the general matrix Π with

$$\Pi = AB^\top + \begin{pmatrix} 0 & 0 \\ 0 & \alpha_{22} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix}\begin{pmatrix} I_r & \beta^\top \\ 0 & I_{k-r} \end{pmatrix}. \tag{11}$$

The null hypothesis now reduces to $\mathrm{rank}(\Pi) = r$, or $\alpha_{22} = 0$. We estimate the model under the null hypothesis using the generalized pseudo log likelihood (10). Then we compute the Hessian $\mathcal{H}$ and the gradient $\mathcal{G}$ of (10) with respect to θ and α_{22} and evaluate these in the estimates under the null hypothesis. Let S_1 be a selection matrix such that $S_1^\top\mathcal{G}$ contains those elements corresponding to the derivatives with respect to the elements in α_{22}, and let S_2 denote the sample estimate of the outer product of gradients. Then the LM-test is given by

$$LM = \left(S_1^\top\mathcal{H}^{-1}\mathcal{G}\right)^\top\left(S_1^\top\mathcal{H}^{-1}S_2\mathcal{H}^{-1}S_1\right)^{-1}\left(S_1^\top\mathcal{H}^{-1}\mathcal{G}\right). \tag{12}$$

Lucas (1996b) contains some explicit formulae for $\mathcal{G}$ and $\mathcal{H}$ and some further simplifications along with the limiting distribution of (12).

To conclude this section, we spend a few lines on what happens if we use the generalized pseudo maximum likelihood estimator in case there are no outliers. Suppose that there are no outliers and that ε_t is normally distributed. Then the proposed procedure suffers from two types of efficiency losses. First, by using a Student $t(5)$ pseudo likelihood instead of the optimal Gaussian likelihood, one looses 5 to 7 per cent of efficiency. This efficiency loss under normality is counterbalanced by a reduced sensitivity and an increased efficiency under deviations from the normality assumption. Simulations in the references mentioned earlier reveal that the reduction in sensitivity and the gain in efficiency under nonnormality generally outweigh the efficiency loss under normality. The second source of efficiency loss comes from the use of the GM weights w_{xt}. In fact, the generalized pseudo likelihood follows the Mallows' specification of GM estimators, see Hampel et al. (1986), meaning that outlying values of Z_t receive a small weight independent of whether the corresponding residual ε_t is small or large. Other weighting schemes are possible, see Hampel et al. (1986), but the Mallows' specification generally has a more stable convergence behavior. The efficiency loss due to using Mallows' weighting scheme can, however, be substantial if only innovation outliers occur. Given the patters in Figures 1 to 9, it seems likely that not only innovation outliers, but also additive outliers are present in the data. Therefore, the use of Mallows' weighting scheme seems appropriate for protection to outliers in the present context of scanning data.

4. EMPIRICAL RESULTS

In this section we discuss the results of determining the cointegration rank for the three-variate system containing m_t, d_t, and p_t. In the next subsection we give some more detailed results for two of the nine brands.

4.1. *Determining the cointegration rank for all data*

In the practical application of cointegration tests, there are various decisions to be made. The first decision concerns the order of the vector autoregression in (5). Since we wish to abstain from a discussion of the impact of irregular observations and of robust techniques on order selection, we decide to calculate our cointegration tests for VAR orders 1, 2, 3, and 4. The second decision concerns the significance level used for the testing procedures. Since the critical values of our test statistics are generated by simulation, see Lucas (1996b), there may be some imprecision in the inference procedure. Therefore, we decide the evaluate our cointegration results at both the 5% and the 10% level. The third decision concerns the inclusion of deterministic components in our regression models. It is well known that redundant deterministics lead to a reduction in power, and that erroneously neglected deterministics lead to severe size distortions. To be on the safe side, we calculate our statistics using both models that include and exclude deterministic linear time trends, i.e., models with $\gamma = 0$ and models with $\gamma \neq 0$ in (5). Note that for the models with linear trends this requires the limiting results in Lucas (1996a) in order to generalize the inference procedure put forward in Lucas (1996b). Finally, as mentioned in Section 3, for our robust approach we decide to follow the suggestions in Franses and Lucas (1996), where through simulations it is found that a Student t likelihood with 5 degrees of freedom yields useful results.

In Table I we summarize our results for the nine brands by presenting the estimated cointegration ranks. Rank 3 implies that all three variables are stationary, rank 0 implies that there is no cointegration amongst the variables (and hence that VAR models in first differences should be considered), and ranks 1 and 2 imply that there is cointegration. It is immediately clear from the columns of Table I that the estimated cointegration rank can vary across robust and non-robust methods. This result holds true for different choices of the significance level and for the two treatments of the deterministic trend. Consider, for example, the first panel containing the results for the case without a trend ($\gamma = 0$ in (5)) at the 5% level. This panel contains 36 pairs of cointegration ranks, four for each brand. The first number of these pairs corresponds to the cointegration rank found with the Gaussian cointegration LM-test, while the second number gives the rank found with the Student $t(5)$ approach. We find that the rank found using the Gaussian method exceeds that of the robust approach in 12 cases, while these ranks are equal in 16 cases. Similarly, one can draw conclusions for the other three panels, for which one can observe that there are many occasions for which the cointegration ranks differ. Whether these differences are due to additive outliers, which may lead to spurious stationarity in a Gaussian framework as shown in Franses and Haldrup (1994), or to innovative outliers and mean shifts, which lead to spurious unit roots (or too small a rank) as shown in Perron and Vogelsang (1992), can now be investigated for particular examples.

4.2. *Some selected results for specific brands*

To save space, we give some more details for only two brands. We decide to consider brands A2 and C2 because the differences between the cointegration ranks seem quite large for these brands, as can be observed from Table I. We present some estimation results for the case where the models contain no deterministic trends, and where we set the significance level at 10% to estimate the cointegration rank. For A2 we then have a rank 2 for the robust method for all VAR orders, while for the Gaussian method the rank is 2 for a VAR(1) and a VAR(2) and zero otherwise. Hence, using the Gaussian method for a VAR(3) and a VAR(4) model one may conclude that there is no cointegration at all. For brand C2 we find for our robust method that the cointegration rank equals 1 for VARs of order 2, 3 and 4, and equals 3 when a VAR(1) is considered. With the Gaussian method we find ranks 3, 2, 1, and 1 for VAR orders 1 through 4.

The obtain our final empirical models, we sequentially delete all regressors which have parameters that are insignificant at the 5% level, starting with the parameters that are least significant. Hence, all parameters to be reported below are significant at the 5% level. We confine our discussion to the long-run parameters linking market share to distribution, and to the short-run parameters concerning the effects of deviations from long-run equilibria between m_t and d_t. In fact, when there are two cointegration relations for the brands A2 and C2, we find that one of the stable relations corresponds to the price variable. Hence for these two brands we find that prices are stationary. Needless to say, that this result does not necessarily carry over to the other brands.

Let β denote the long-run elasticity of distribution on market share. Furthermore, let α_1 and α_2 denote the short-run responses of market share and relative distribution to disequilibrium errors in the equilibrium relation linking market share to distribution. The final estimates for these parameters are presented in Table II.

We start with a discussion of the Gaussian results. For A2 we obtain for both the VAR(1) and VAR(2) regression models that β is estimated as 1.69. The corresponding α_1 parameters are estimated as -0.19 and -0.16, respectively, while the α_2 parameters are equal to 0.33 and 0.20. Using our robust methods, we find for the corresponding VAR orders that β is estimated as 1.94 and 1.90, while the short-run adjustment parameters are -0.13 and -0.10 for α_1 and 0.23 and 0.17 for α_2. For the VAR(3) and VAR(4), the Gaussian method does not find any cointegration. By contrast, the robust method still finds two cointegration relations and yields 2.06 and 2.07 for β, 0.00 for α_1, and 0.12 for α_2 for the VAR(3) and VAR(4), respectively. Hence, the long-run parameter appears to be estimated at a constant value using the robust method, while the Gaussian method only results in such constant values for a smaller set of models. If we compare the different results for A2, we find that the long-run elasticity consistently takes larger values when using the robust method, while the absolute values of the short-run correction parameters tend to be smaller. Since both the robust and nonrobust approach yield significant equilibrium correction, we can interpret the findings for A2 as indicating that there are significant long-run as well as short-run effects.

For brand C2, we find that the Gaussian method results in rank $r = 2$

for a VAR(2) and a rank $r = 1$ for a VAR(3) and a VAR(4). In the first case, we again find that the price variable is stationary. The β parameter in the VAR(2) equals 2.10, which seems to match quite well with the results for brand A2. This also applies to the short-run parameters, which are -0.15 and 0.17 for α_1 and α_2, respectively. For a VAR(3) and a VAR(4), however, only one cointegration relation is found. For the VAR(3) this results in an estimate of β of 3.55, while α_1 equals 0.00 and α_2 equals 0.23. For the VAR(4) we obtain $\beta = 4.79$, $\alpha_1 = 0.12$, and $\alpha_2 = 0.17$, such that the model does not equilibrium-correct because $\alpha_1 > 0$. Hence, for the Gaussian case, the long-run elasticity is not constant over the VAR orders. This is in contrast to the robust approach, for which we estimate the rank as 1 for the VAR(2), VAR(3), and VAR(4). The long-run elasticity β now equals 4.00, 4.26, and 4.75 for these models. Thus, the apparent correspondence between A2 and C2 disappears. Furthermore, for a VAR(2) and a VAR(3), α_1 can be set to zero, while for the VAR(4) we have $\alpha_1 = 0.05$. This last value again indicates that the VAR(4) model is not equilibrium correcting, although the value 0.05 is much smaller than the value of α_1 estimated with the nonrobust approach. Using the robust approach, we get the values 0.19, 0.20, and 0.20, for α_2 in the VAR(2), VAR(3), and VAR(4), respectively. This parameter is thus very stable over the different orders of the VAR. For C2 we conclude that there are only significant long-run effects for distribution, since α_2 is unequal to zero while α_1 is zero.

The effects of irregular observations on our econometric models can be highlighted by graphs of the weights of the observations, see also Franses and Lucas (1996). In Figures 10 and 11 we give these weights for brand A2 and C2, respectively. The observations marked with an inverted triangle are data points that receive a weight that is significantly below a certain threshold. As is obvious from these two graphs, there are many significant irregular data points. This emphasizes the usefulness of robust estimation methods when analyzing weekly scanning data.

5. CONCLUDING REMARKS

In this paper we advocated the application of robust cointegration analysis to investigate the correlations between market share and distribution in weekly observed scanning data. Typically, such highly disaggregated data can be contaminated by many irregular observations, as could be observed from the graphs in this paper for nine brands of a fast-moving consumer product. Furthermore, market share and distribution time series may display trending patterns, and hence one has to rely on cointegration techniques for proper statistical inference. We showed that the results for the long-run properties of the marketing time series variables can differ to a large extent across robust and nonrobust approaches. In turn, these differences can yield different outcomes for the long-run and short-run effects of distribution on market share and vice versa. Since these outcomes have a marketing interpretation, we recommend the marketing researcher to apply both approaches. If the differences are small, one can be confident with the empirical results. Additionally, any empirical generalizations obtained from these results may become useful for further research. When a robust analysis yields results that are markedly different from results obtained by a Gaussian analysis, however, we feel that one should be very careful with an interpretation of the findings. More so, we believe that empirical marketing generalizations cannot be stable if they can be shown to depend largely on irregular observations.

REFERENCES

Bass,Frank M. and Jerry Wind (1995, eds.): Empirical Generalizations in Marketing, Special Issue of *Marketing Science*, Linthicum: The Institute for Operations Research and the Management Sciences.

Farris, Paul, James Olver, and Cornelis De Kluyver (1989): "The Relationship Between Distribution and Market Share," *Marketing Science* 8, 107-128.

Franses, Philip Hans, and Niels Haldrup (1994): "The effects of additive outliers on tests for unit roots and cointegration," *Journal of Business and Economic Statistics* 12, 471–478.

Franses, Philip Hans, and André Lucas (1995): Outlier detection in cointegration," Mimeo, Department of Financial Sector Management, Free University Amsterdam (Revision submitted for publication).

Gouriéroux, Christian, Alain Monfort, and Alain Trognon (1984): "Pseudo maximum likelihood methods: theory," *Econometrica* 52, 681–700.

Hampel, F.R., E.M. Ronchetti, Peter J. Rousseeuw, and W.A. Stahel (1986): *Robust Statistics. The Approach Based on Influence Functions.* New York: Wiley.

Hoek, H., A. Lucas, and H.K. van Dijk (1995): "Classical and Bayesian aspects of robust unit root inference," *Journal of Econometrics* 69, 27–59.

Huber, P.J. (1981): *Robust Statistics.* New York: Wiley.

Johansen, Søren (1988): "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control* 12, 231–254.

Johansen, Søren (1991): "Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models," *Econometrica* 59, 1551–1580.

Johansen, S. (1994): "The role of the constant and linear terms in cointegration analysis of nonstationary variables," *Econometric Reviews* 13, 205–229.

Kleibergen, Frank R., and Herman K. van Dijk (1994): "Direct cointegration testing in error correction models," *Journal of Econometrics* 63, 61–103.

Lucas, André (1996a): *Outlier Robust Unit Root Analysis*. Ph.D. Thesis, Amsterdam: Thesis Publishers.

Lucas, André (1996b): Inference on cointegrating ranks using pseudo LM tests, Mimeo, Department of Financial Sector Management, Free University Amsterdam.

Lucas, André (1997): "Cointegration testing using pseudo likelihood ratio tests," *Econometric Theory* 13, (forthcoming).

Martin, R.D. (1979): "Robust estimation for time series autoregressions," in *Robustness in statistics*, R.L. Launer and G. Wilkinson (eds.), New York: Academic Press, 147–176.

Martin, R.D. (1981): "Robust methods for time series," in *Applied time series analysis II*, David F. Findley (ed.), New York: Academic Press, 683–759.

Perron, Pierre and Tim J. Vogelsang (1992): "Nonstationarity and level shifts with an application to purchasing power parity," *Journal of Business and Economic Statistics* 10 , 301–320.

White, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica* 50, 1–25.

TABLE I

COINTEGRATION RANKS FOR VARIOUS VAR ORDERS USING ROBUST AND NONROBUST TESTS

p	A1		A2		A3		B1		B2		B3		C1		C2		C3	
	G	t	G	t	G	t	G	t	G	t	G	t	G	t	G	t	G	t
	No trend, 5%																	
1	3	3	2	2	3	3	3	3	2	3	2	3	3	3	3	3	3	3
2	2	3	2	2	2	2	2	2	2	1	2	1	3	3	2	1	3	3
3	2	3	0	1	2	2	2	2	2	1	2	1	2	2	1	0	3	2
4	2	3	0	1	1	2	2	2	1	0	2	1	2	0	1	0	3	2
	No trend, 10%																	
1	3	3	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2	3	3	2	2	3	2	3	3	2	1	2	1	3	3	2	1	3	3
3	3	3	0	2	2	2	3	3	2	1	2	1	3	2	1	1	3	2
4	3	3	0	2	2	2	2	3	2	0	2	1	3	2	1	1	3	2
	Trend, 5%																	
1	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2	2	3	1	3	3	3	3	1	2	1	2	3	3	3	3	2	3	2
3	2	3	0	1	2	1	3	1	1	1	2	3	3	3	1	1	2	2
4	2	3	0	1	2	2	2	1	1	0	2	1	3	3	1	1	2	2
	Trend, 10%																	
1	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2	2	3	2	3	3	3	3	1	3	2	2	3	3	3	3	2	3	2
3	2	3	0	2	3	2	3	3	2	2	2	3	3	3	2	1	3	2
4	3	3	0	1	2	2	2	1	2	0	2	3	3	3	1	1	2	2

NOTE: The table contains the estimated cointegration ranks for the different brands using the Gaussian-based LM-test (G) and the Student $t(5)$-based LM-test (t). The VAR model of order p is estimated with an unrestricted constant (No trend) and an unrestricted trend (Trend), respectively. Cointegration testing procedures using a 5% and a 10% level are employed.

TABLE II

FINAL ESTIMATES OF LONG-RUN ELASTICITIES AND SHORT-RUN DISEQUILIBRIUM RESPONSES FOR A2 AND C2

p	A2						C2					
	G			t			G			t		
	β	α_1	α_2	β	α_1	α_2	β	α_1	α_2	β	α_1	α_2
1	1.69	-0.19	0.33	1.94	-0.13	0.23	—	—	—	—	—	—
2	1.69	-0.16	0.20	1.90	-0.10	0.17	2.10	-0.15	0.17	4.00	0.00	0.19
3	—	—	—	2.06	0.00	0.12	3.55	0.00	0.23	4.26	0.00	0.20
4	—	—	—	2.07	0.00	0.12	4.79	0.12	0.17	4.75	0.05	0.20

NOTE: The table contains the final estimates of the long-run elasticity β of distribution on market share, and the short run responses of market share (α_1) and distribution (α_2) to disequilibrium errors in the market share equilibrium relation, i.e., $m_t = \beta d_t + \tilde{\beta} p_t$. The estimates are for various orders p of the VAR model after sequential elimination of all regressors that were insignificant at the 5% level using the Gaussian (G) pseudo maximum likelihood estimator or the Student $t(5)$ (t) generalized pseudo maximum likelihood estimator. The appropriate cointegration rank for each value of p and for each brand is taken from Table I. Entries in the table are indicated as missing if the VAR is either completely stationary ($r = 3$) or nonstationary ($r = 0$).

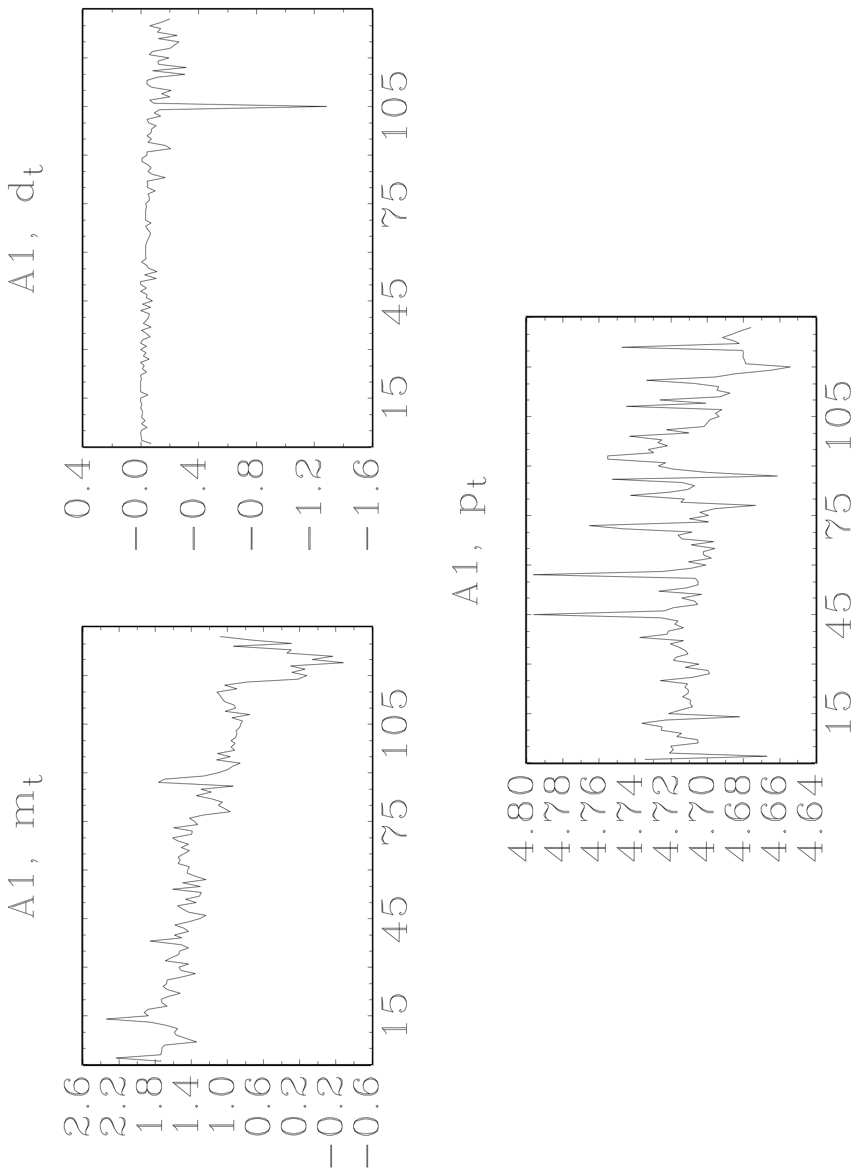

Figure 1.— m_t, d_t, p_t, and ca_t, for brand A1

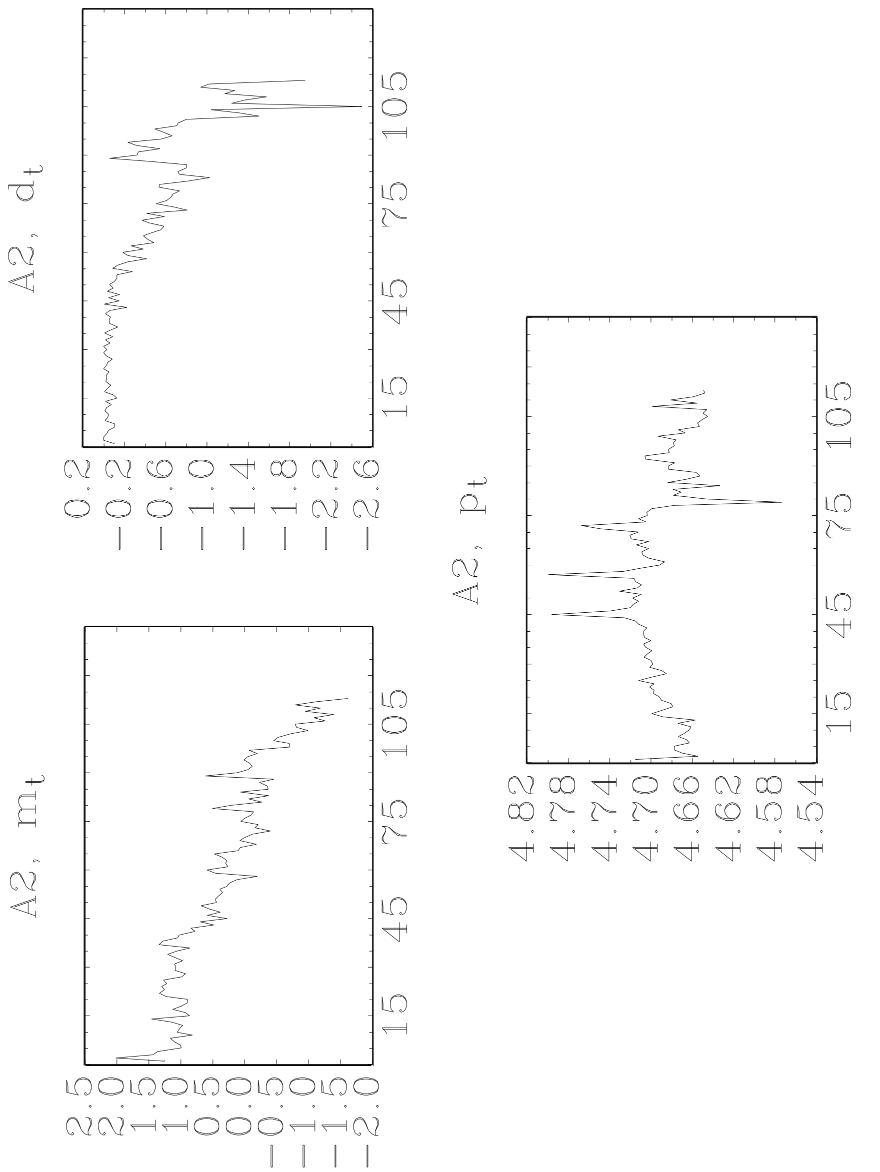

Figure 2.— m_t, d_t, p_t, and ca_t, for brand A2

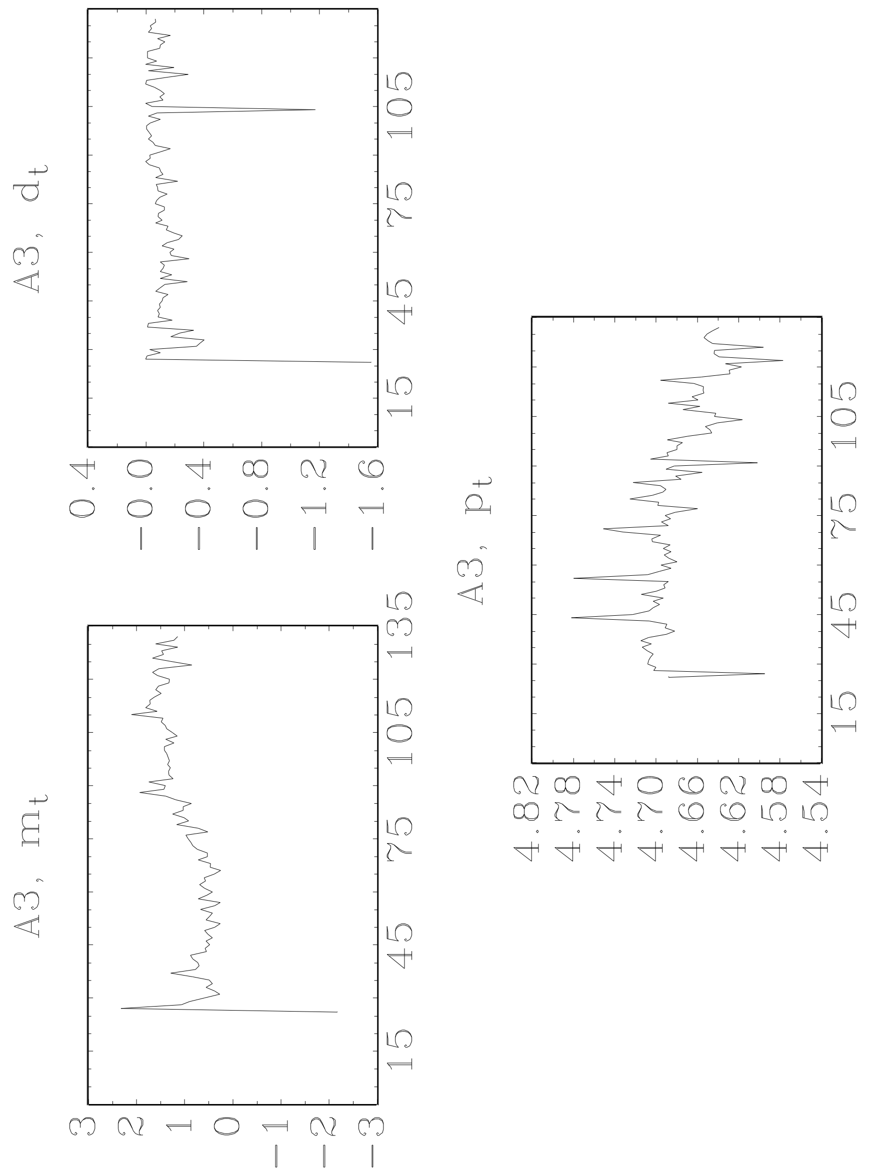

Figure 3.— m_t, d_t, p_t, and ca_t, for brand A3

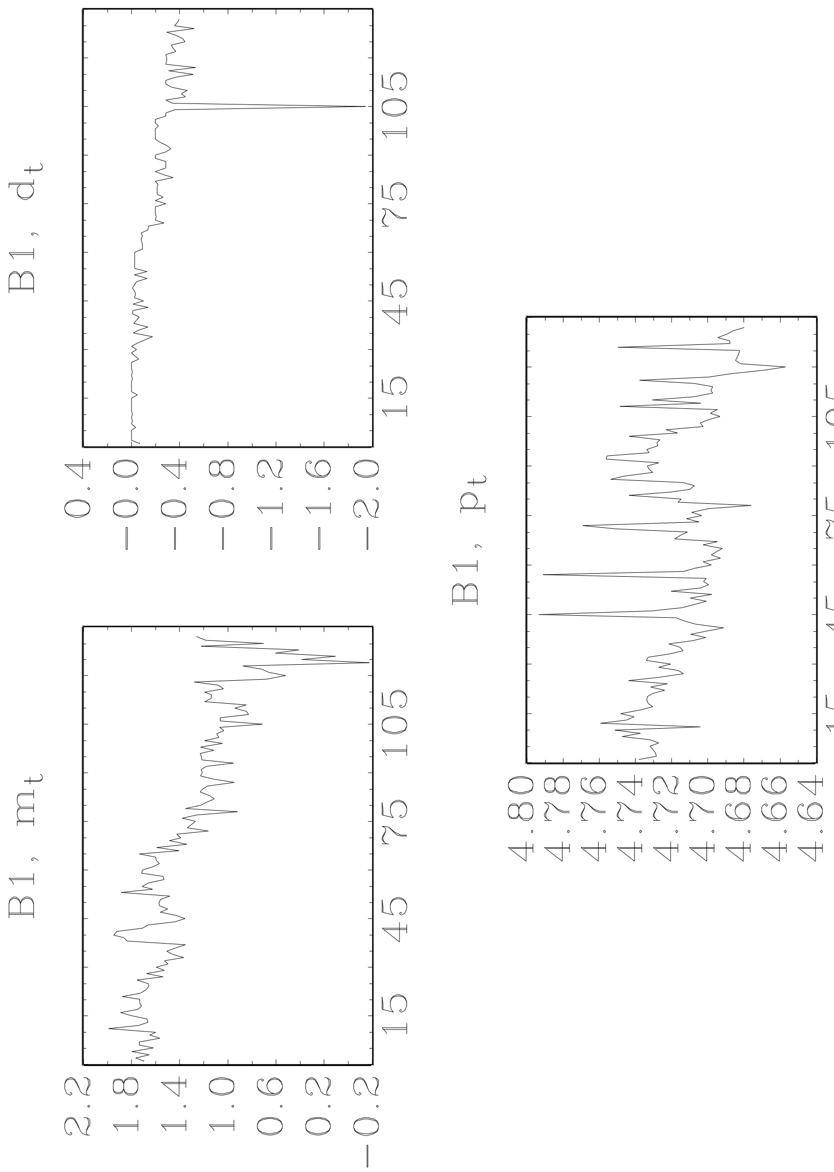

Figure 4.— m_t, d_t, p_t, and ca_t, for brand B1

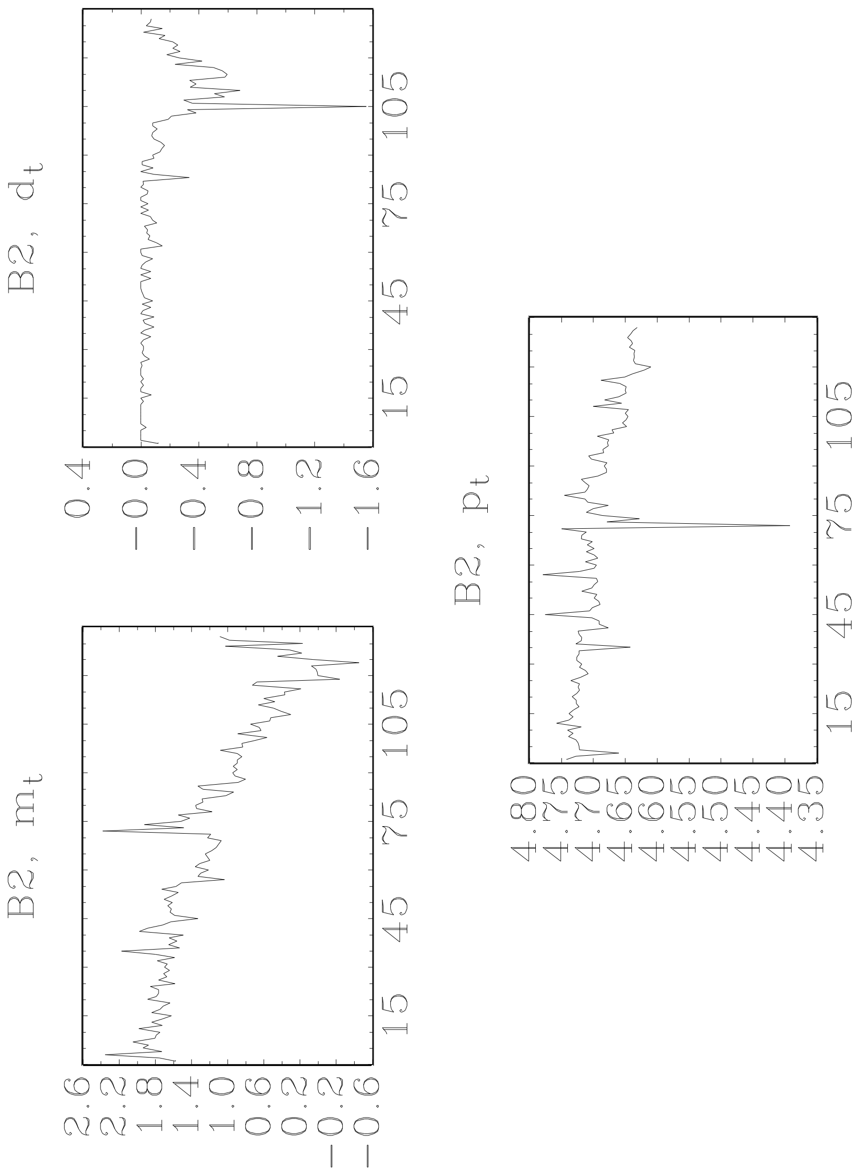

Figure 5.— m_t, d_t, p_t, and ca_t, for brand B2

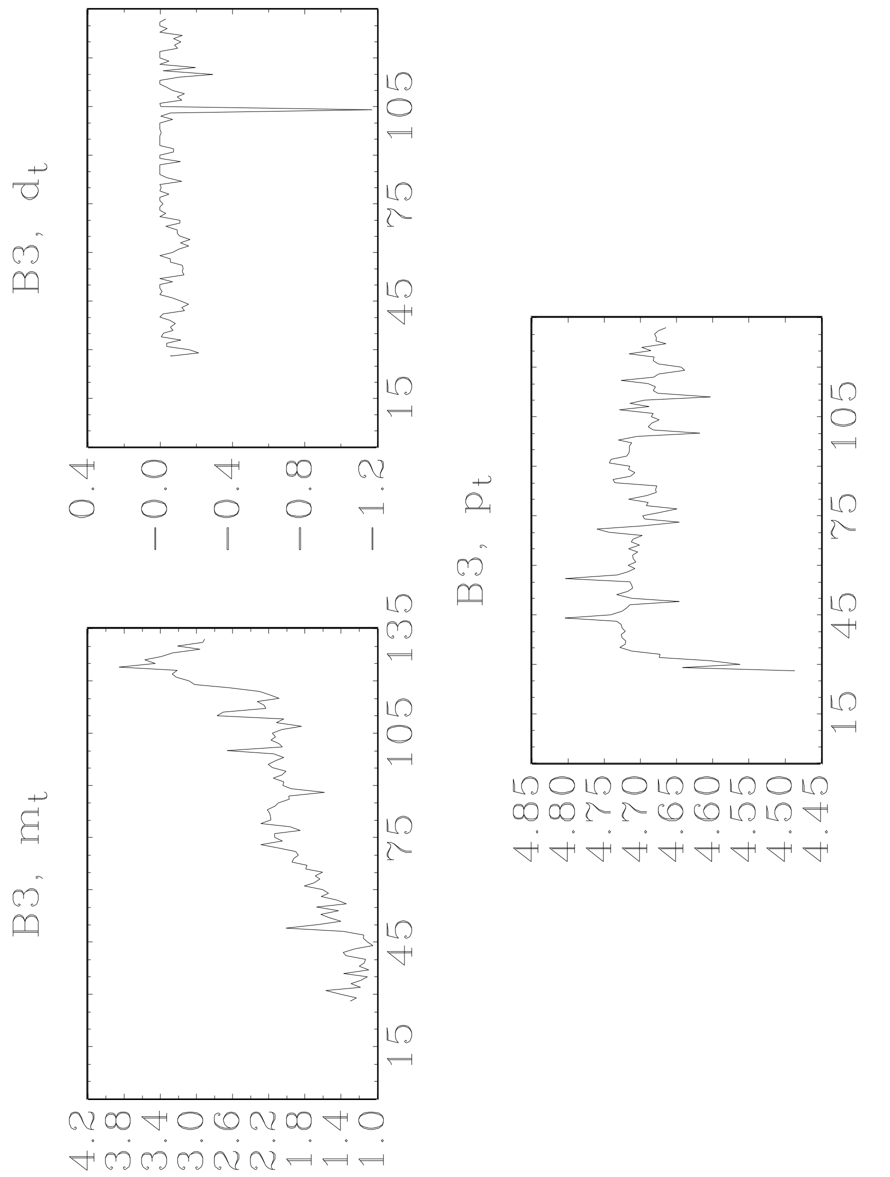

Figure 6.— m_t, d_t, p_t, and ca_t, for brand B3

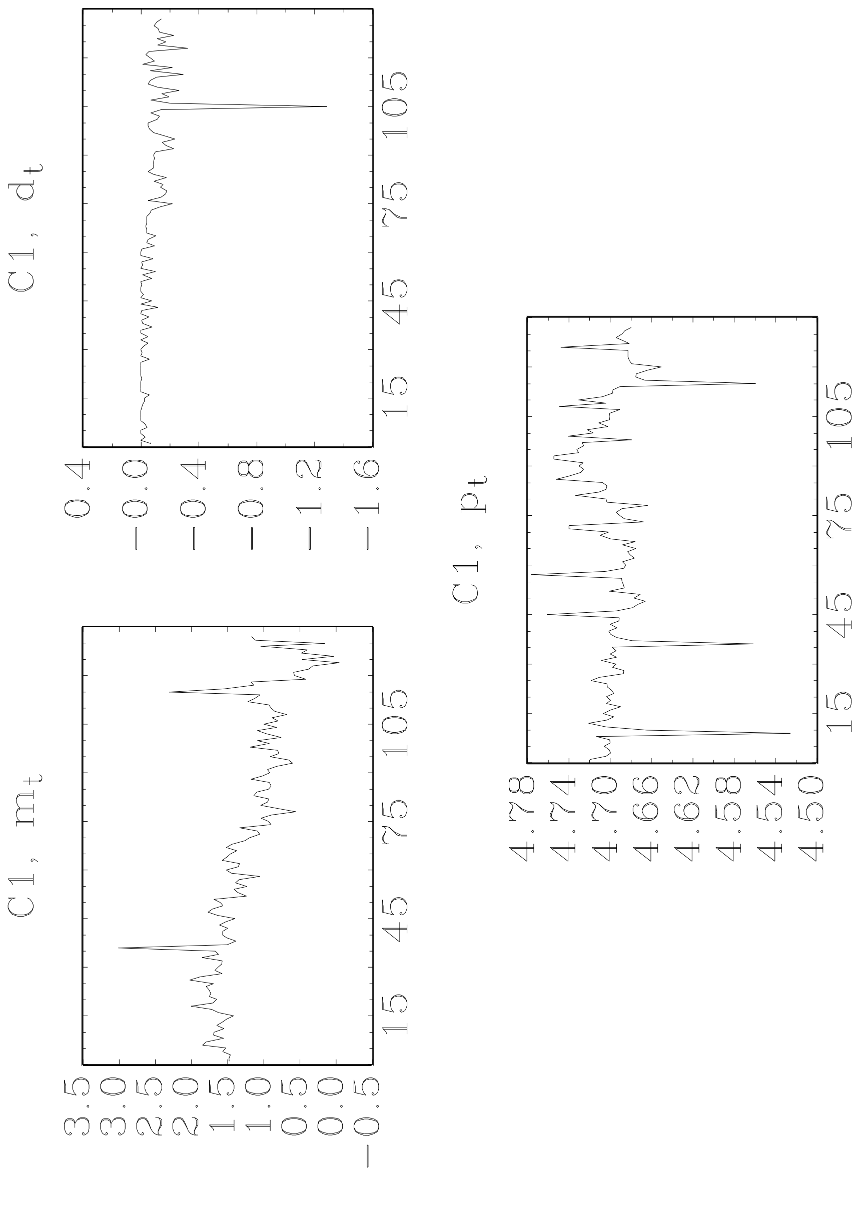

Figure 7.— m_t, d_t, p_t, and ca_t, for brand C1

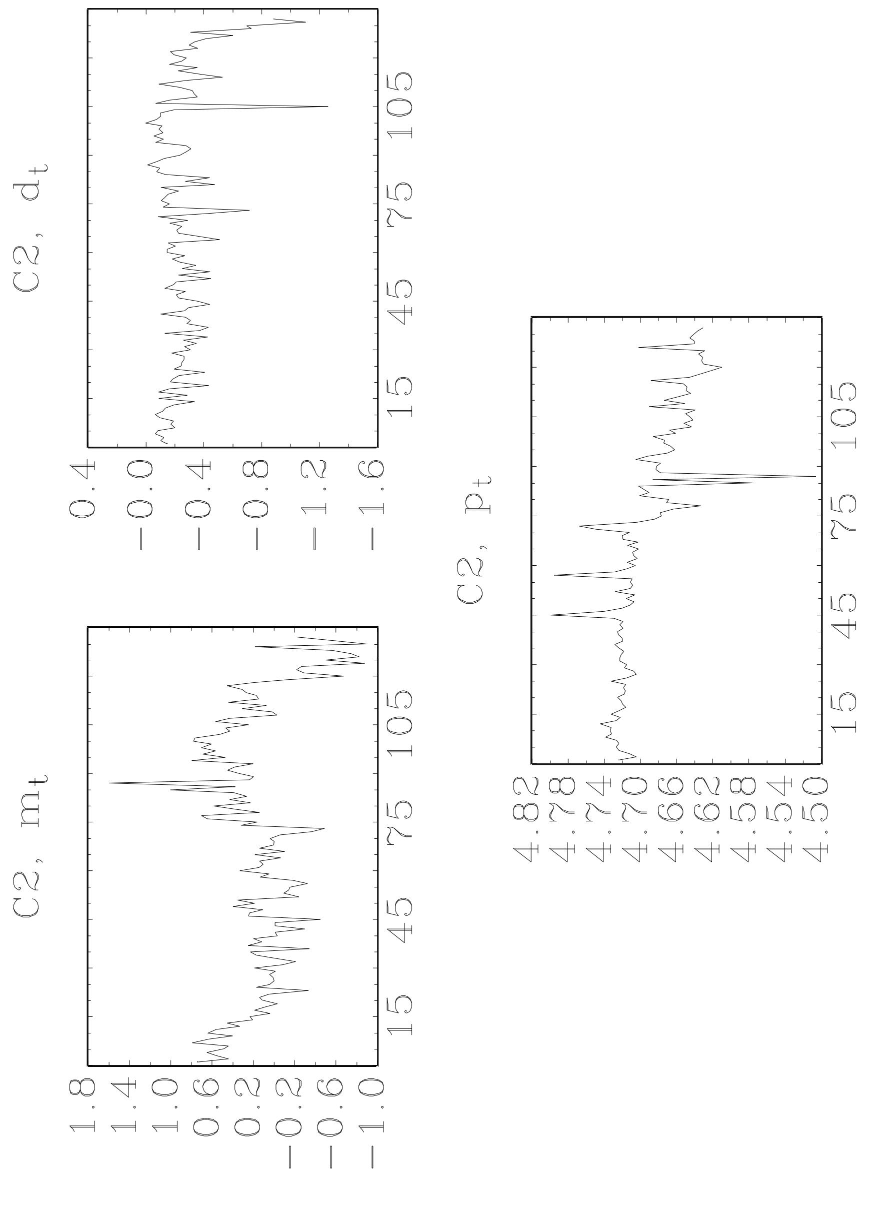

Figure 8.— m_t, d_t, p_t, and ca_t, for brand C2

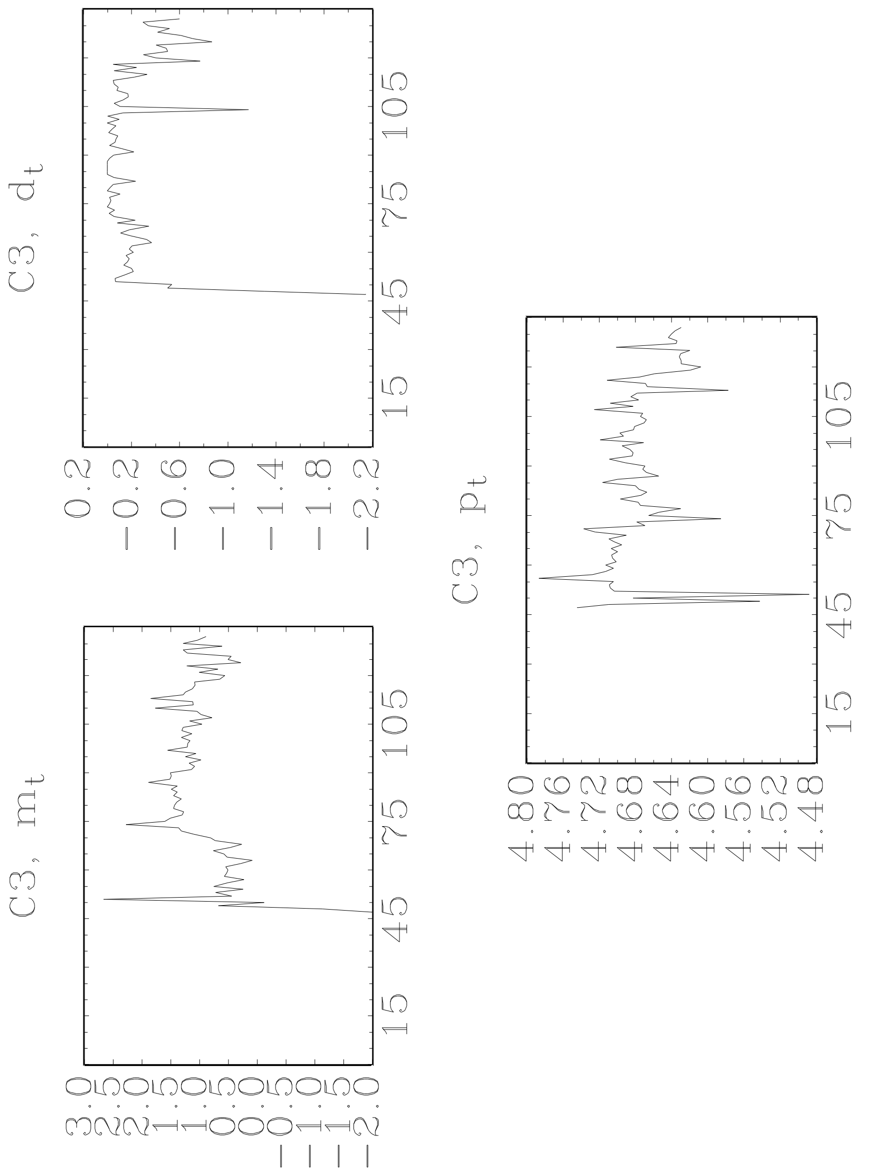

Figure 9.— m_t, d_t, p_t, and ca_t, for brand C3

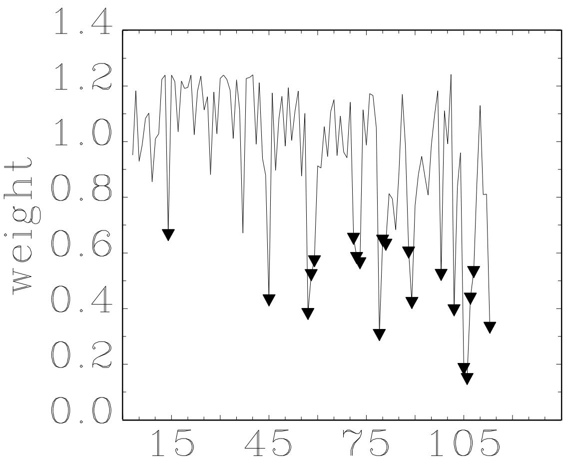

Figure 10.— Weights for VAR($p = 2, r = 2$) for brand A2

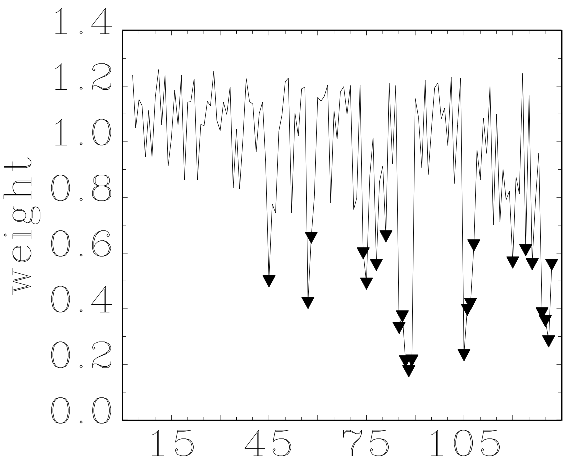

Figure 11.— Weights for VAR($p = 2, r = 1$) for brand C2