*empirical economics, Vol. 2, Issue 4, page 213–223. Physica-Verlag, Vienna.*

# Aggregated Dynamic Demand Equations for Specialistic-outpatient Medical Care

## (Estimated from a Time-Series of Cross-Sections)

By *J. van der Gaag, B.M.S. van Praag, F.F.H. Rutten, W. van de Ven*, Leiden[1])

*Abstract*: In this paper a dynamic model is presented which describes the development of the demand for specialistic medical care in The Netherlands, during the period 1960–1972.
The "regionally correlated, time-wise auto-regressive" model is consistently estimated from a time-series of cross-sections, using a modified Aitken estimator.

The dependent variables are the number of publicly insured patients referred from general care to specialistic care, and the amount of care consumed per patient referred.

As independent variables we took demographic factors, the supply of different levels of medical care and the insurance system. The estimation results show a.o. important substitution possibilities between general and specialistic care, and a significant influence of supply and supply-related variables on the demand for specialistic care.

## 1. Introduction

In the last decade the health care sector has become more and more a field of interest for economic and econometric researchers. Their interest is mainly stimulated by the tremendous increase of the cost of health care which can be observed in almost every country. In The Netherlands for instance about 4,5 % of GNP was spent in the Health Care Sector in 1962. For 1972 the figure is 7,2 % and the estimate for 1980 is about 12 % of GNP.

The idea that "the more the better" is no longer true in the Health Care Sector becomes accepted. But before measures can be taken to stop the cost increase more insight is needed into the structure of the sector.

From an economic point of view two particular features are of interest: the ignorance of the patient with the medical market [see *Arrow*] and the general lack of a (direct) price, due to insurance coverage. The latter is especially important in The Netherlands, where a good 70 % of the population is compulsorily insured against all medical expenses under a Public Health Insurance scheme. It is to this group that we will pay attention in this paper.

---

[1]) Prof. *Bernard van Praag*, Rijksuniversiteit te Leiden, Economic Institute, Hugo de Grootstraat 32, NL–Leiden.

The Dutch practice is that the patient consults a General Practitioner (*GP*) for his complaints. The *GP* decides whether the patient needs a treatment and, if so, whether he treats the patient himself or whether the patient must be referred to a specialist. The *number of referrals* per year per hundred publicly insured is one of our dependent variables. It is a key-variable in the Dutch health care sector, because by far the largest part of total costs is made (after a referral) in the second and third "echelon" (i.e. specialistic outpatient- resp. inpatient-care).

In the next section we will define and discuss a number of variables which, as we think, influence the demand for specialistic-outpatient care (also to be defined).

Then the formal model will be presented with reference to the estimation problems. The estimation results are presented in section 3. Finally the findings are discussed and summarized in section 4.

## 2. The Demand for Specialistic-Outpatient Care

### 2.1 Definition

As stated in the introduction we will restrict ourselves to the publicly insured part of the Dutch population. In 1960 the number of referrals per hundred insured was 33.5. In 1972 it was 46.1. The number of referrals can be seen as the number of people asking for specialistic care. For each person referred to him, the specialist gets a fixed amount. For this amount he treats the patient during a month. The amount does not depend on the number of doctor/patient contacts during that month. If the treatment is continued after a month, the specialist writes a so-called "continuation card", which is also worth a fixed amount[2]) for which another month treatment is given.

In this paper we define the sum of the referrals and continuation cards as the demand for specialistic-outpatient care.[3])[4]) Because the decision whether care is needed is taken by the *GP* for referrals and by the specialist for the continuation of care both parts of medical consumption will be treated separately.

It is useful to consider the physician as an *agent* acting for or his patient [compare *Feldstein*]. The patient enters the medical channel with only a vague notion about the exact type, quantity and quality of care he needs. The physician translates the patient's general demand for care into a specific use of medical resources. As long as only characteristics related to the patient are involved we are not prevented from estimating *demand* equations.

However as soon as the physician's decision (or better: advice) is influenced by, for instance, the availability of alternative types of care, a pure *demand* equation is hard to identify (if at all). Therefore the equations below can best be looked at as "partial

---

[2]) Half of that referral. The number of continuation cards per hundred insured was 27.4 in 1960 and 49.5 in 1972.

[3]) Note that *X*-rays, small operations, prescriptions etc., which are paid separately (by the insurance) are not in this definition.

[4]) Note that we use the word "demand" in total absence of any price.

reduced-form equations" explaining the utilization of health care facilities as the result of a badly defined demand for care (by the patient) *and* a specification of this demand (by the physician) influenced by supply-related variables.[5])

## 2.2 The Independent Variables

With respect to the number of referrals it is important to observe that the *GP* is paid according to a capitation system, i.e. he gets a fixed amount per year for each publicly insured patient "on his list", regardless of the amount of care given. At the same time one must notice that the number of referrals is, in a way, a measure of the number of cases the *GP* is not able to treat. So one could say that the decision to send a patient to a specialist reflects for the *GP* a choice between leisure and quality of work;[6]) this suggests that the number of referrals per thousand will be influenced by the number of patients in his practice: the busier the practice the sooner the patient is referred to a specialist. Therefore the number of *GP's* per 1000 population[7]) is added as an explanatory variable.

Of course things are more complicated than described above. For instance the character of a practice is not fully described by its size; a second feature is the age-sex-structure of the population. Because we are dealing with rather large areas (see sub-section 2.4) the number of women per thousand men shows very little variance, so we decided not to include this variable in the analysis. The age-structure is represented by the percentage of *old-aged* (65+).

A more severe complication is the dichotomy between publicly and privately insured patients, especially because the latter pay the *GP* a fee for service. So the allocation of time is a threefold choice for the *GP*, viz., between quality of work, leisure and/or additional income (from the treatment of private patients). A full treatment of this problem is only possible, if data about the number of patient/doctor contacts and referrals of all patients (publicly and privately insured) are available, which is not the case at the moment.

However, if the treatment of a patient by the *GP* does not depend on the remuneration system [but see for instance *Roemer*] and if the two insured groups do not differ significantly with respect to their demand for general health care, the number of referrals per hundred publicly insured cannot depend on the percentage of publicly insured patients in the practice. We add the percentage of *publicly insured patients* as an explanatory variable to test this hypothesis.

Apart from the "demand-creating" character of health-care supply [see for instance *Fuchs/Kramer*] it is clear that the *GP* will have to treat the patient himself, if no or too little specialistic care is available. The availability is measured by the number of *specialists* per 1000 population and by the *population density*, as a measure of distance.

---

[5]) Note that we use the word "demand" in total absence of any price.
[6]) For a given income.
[7]) The reciprocal of the size of the practice.

Because applications of advanced medical technology are generally to be found in urban areas, the – monotonously increasing – population density must also be interpreted as a trend variable reflecting increasing medical possibilities, changing medical insight, etc.[8])

Finally the number of beds in *university hospitals* is added as a measure of specialistic care available in university hospitals because the number of referrals to specialists working in university hospitals is not included in our dependent variable.

For the number of *continuation* cards per referral, i.e. the amount of care given to the *referred* patient, we assume that the size of the practice of the specialist is of influence, and the size of the practice of the *GP* (as a measure of the availability of general care, for the follow-up treatment).
The percentage old-aged, the population density and the percentage of publicly insured patients are added for analogous reasons as with respect to the number of referrals.[9])
It is plausible that the more difficult cases are sent to specialists in university hospitals, if available. Consequently, one can expect a negative influence of this variable on the number of continuation cards (for the less difficult referrals, on which less care is given). Finally we assume that if the number of referrals per insured is relatively high, the average amount of care "consumed" by a *referred* patient will be relatively low. Therefore the number of referrals per 100 publicly insured is added as an explanatory variable.

## 2.3 A Dynamic Specification

It is plausible to assume that the decision makers in the medical field, the *GP* and the specialist, are more or less "persistent" in their "habits" [e.g. *Brown*]. They have developed a way of working with which they feel comfortable. These habits will only change slowly under influence of exogenous variables.
Besides this argument in favour of a dynamic specification of the equations, there is another:
If for instance the number of *GP*'s increases with, say 10 %, we measure that the (average) size of the practice decreases about 10 %, but in reality we are only dealing with a few additional small practices. Most existing practices will decrease in size only slowly toward the average, so the effect on the dependent variables will also appear "in time".

These arguments (and others for the other exogenous variables) lead to a specification in which the lagged dependent variable is included as predetermined at the right-hand side of the equations. So, to avoid major difficulties, we assume the average time-lag to be the same for all variables.

---

[8]) The equations are estimated from a time-series of cross-sections. The population density is the only "trend variable" available that differs between the regions. Experiments with an income variable (the same for all regions) and a time-index did not yield satisfactory results.

[9]) The specialist is also paid a fee for service by the private patient.

### 2.4 The Model

The data stem from a large variety of sources. If necessary they have been aggregated per province. The Netherlands count eleven provinces ($N = 11$). The data are collected for the period 1960–1972 ($T = 13$).

The model looks as follows:

$$y_{1it} = x'_{it}\beta_1 + \lambda_1 y_{1i,t-1} + u_{1it}$$

$$y_{2it} = x'_{it}\beta_2 + \gamma y_{1it} + \lambda_2 y_{2i,t-1} + u_{2it} \qquad \begin{array}{l} i = 1, N \\ t = 2, T \end{array}$$

with:

$y_{1it}$: the number of referrals per hundred publicly insured, in province $i$, in year $t$.

$y_{2it}$: the number of continuation cards per referral

$x'_{it}$ is a vector of the following predetermined variables (a constant term is added):[10])

$GP_{it}$ : the number of general practitioners per 1,000 population

$SP_{it}$ : the number of specialists per 1,000 population

$AGE_{it}$ : the percentage of old-aged (65+)

$DENS_{it}$ : the population density (number of inhabitants per km$^2$)

$PIN_{it}$ : the percentage of publicly insured patients

$UN_{it}$ : the number of beds in university hospitals per 1,000 population.

The vectors $\beta_1$ and $\beta_2$ and the coefficients $\lambda_1$, $\lambda_2$ and $\gamma$ have to be estimated; $u_{1it}$ and $u_{2it}$ are independent disturbances. Because the model is recursive, the equations can be estimated one by one.[11]) We will deal with the estimation problems in the next section.

### 2.5 Estimation Problems

Let

$$y_{it} = x'_{it}\beta + \lambda y_{i,t-1} + u_{it}$$

be the model to be estimated. We assume[12])

---

[10]) All variables $y$ and $x$ are logarithmic transforms.

[11]) Actually we are dealing with two sets of "seemingly unrelated regressions" [*Zellner; Parks*] for which per set the coefficients are the same over the equations.

[12]) An error component specification as for instance used by *Balestra/Nerlove* [1966] implies a rather peculiar regional correlation and autoregressive structure [see *Kmenta*, p. 515], which makes such a specification less attractive in our case; though the estimation results of the error-component-model (available from the authors) and our error specification are in close agreement.

$$E(u_{it}^2) = \sigma_{ii}$$

$$E(u_{it}, u_{jt}) = \sigma_{ij}$$

$$u_{it} = \rho_i u_{i,t-1} + \epsilon_{it} \qquad i = 1, N \qquad t = 2, T$$

with $\epsilon_{it}$ a "real" disturbance, independent of $u$.

So besides heteroskedasticity, we also assume inter-region correlation. We cannot circumvent this complication because we are dealing with geographical regions (provinces). The medical facilities in one province are in principle available for all inhabitants of the other provinces (and especially for the adjacent provinces).[13]) Furthermore we assume first-order auto-correlation, which may vary between the regions; the rationale for this is that we cannot rule out a possible influence of variables that change over time, but which are only important for a specific province.[14]) These variables may induce auto-correlation in the disturbances which cannot be assumed to be the same for all provinces.

Now the variance-covariance matrix $\Sigma$ depends on $1/2 \cdot N \cdot (N + 1) + N$ parameters and looks as

$$\Sigma = \begin{bmatrix} \sigma_{11}P_{11} & \sigma_{12}P_{12} & \cdots & \sigma_{1N}P_{1N} \\ \sigma_{21}P_{21} & \sigma_{22}P_{22} & & \\ & & \ddots & \\ \sigma_{N1}P_{N1} & & & \sigma_{NN}P_{NN} \end{bmatrix}$$

with

$$P_{ij} = \begin{bmatrix} 1 & \rho_j & \rho_j^2 & \cdots & \rho_j^{T-1} \\ \rho_i & 1 & \rho_j & & \\ \rho_i^2 & \rho_i & 1 & & \\ & & & \ddots & \\ \rho_i^{T-1} & & & & 1 \end{bmatrix}$$

If a lagged dependent variable is included in a model with auto-correlated errors we get some well-known estimation difficulties. OLS does not yield a consistent estimate of the disturbances in the first stage of an Aitken estimation procedure, so we cannot find a consistent estimate of $\Sigma$.

---

[13]) Moreover some of the data stem from regions that were smaller than provinces. In the aggregation "border-problems" had to be solved, which makes the absence of inter-region correlation unrealistic.

[14]) For instance the vast increase of unemployment of miners in one province during the period considered.

There exists an impressive amount of literature about the estimation of distributed-lag-models [*Griliches; Dhrymes; Sims*]. For our purpose an Instrumental Variable approach, as suggested at first by *Balestra/Nerlove* [1966] in this context, seems appropriate.

*Wallis* [1967] showed that GLS has to be preferred to Three-Pass Least-Squares (3PLS; *Taylor/Wilson* [1964]) and also suggested the instrument to be used if more than one exogenous variable is involved.

Though we know that 3PLS and GLS fall short of asymptotic efficiency [*Grether/Maddala*], very little is known about the small sample properties. *Hatanaka* [1974] proved that the "residual-adjusted Aitken estimator" he derived is asymptotically identical to the maximum likelihood estimator, and moreover, he showed that the two-step Aitken method performed equally well as the residual-adjusted estimator in small samples.

For a situation in which information is available from a time-series of cross-sections *Nerlove* [1971] showed that a step procedure performed even better than maximum likelihood in a small sample. However, *Nerlove* was dealing with an error-component model, without a pure time effect and for relatively small $T$, so his results cannot be directly extended to our more general model specification.

Given these results from the literature we chose the following modification of the Aitken-estimator as appropriate for our purpose:

1. First regress $y_{it}$ on all exogenous variables, without including the lagged dependent variable.
2. Use the estimated $y_{i,t-1}$ in the regression as the lagged dependent (instrumental) variable.
3. Use the consistently estimated disturbances of the second step to estimate $\rho_i$, $(i = 1, N)$ and to apply the usual transformation on $y$ and $x$.
4. Apply OLS on this transformed model to obtain a consistent estimate of the variance-covariance matrix of the disturbances.
5. Estimate the coefficients by GLS on the transformed model. [See *Kmenta* for details]

## 3. Estimation Results

The estimation results are given in Table 1.

The influence of *GP* on the demand for specialistic outpatient care is negative, as expected. We see that if the number of general practitioners per 1,000 population increases, less patients will enter the "second echelon". And the patients who are referred to the specialist will sooner be back in the "first echelon" if many *GP*'s are available.

The availability of specialists has also a positive impact on "demand". Both the number of referrals and the number of continuation cards will rise if *SP* increases.

The well-known fact that the aged people consume more medical care is reflected in the coefficients 0.08 and 0.24 for $y_1$ and $y_2$ respectively. The elasticities with respect to *UN* are also as expected.

Of more interest are the elasticities with respect to *DENS*. If we interpret *DENS* as a trend variable, we see that the general practitioner is more and more inclined to refer a patient to a specialist. A referred patient gets more attention of the specialist each year. Probably the increase in (diagnostic) medical-technical possibilities is responsible for this phenomenon.[15])

We also see that the demand for specialistic care seems to depend on the insurance system.[16]) If the percentage of publicly insured patients increases, the number of referrals per publicly insured increases. In other words, if the publicly insured patients constitute a relatively large group in a practice of a general practitioner they get relatively less attention of the general practitioner (the number of referrals is high). This does not seem to be true for specialistic care: *PIN* does not show a significant influence on $y_2$.[17])

The average time-lag is rather large: for $y_1$ it is about 4 years. The coefficient for the lagged dependent variable $y_2$ is only 0.25. Since $y_2$ is the number of continuation cards *per referral* the second time-lag has to be added to the time-lag of $y_1$.

The vector of long-term elasticities, $b_1^*$, of the number of referrals per insured can be calculated as follows:

$$b_1^* = b_1/(1 - \alpha_1)$$

with

$b_1$: the vector of estimated coefficients of the exogenous variables

$\alpha_1$: the estimated coefficient of the lagged dependent variable $y_{1i,t\text{-}1}$

In an analogous way the long-term elasticities of the number of continuation cards *per insured* ($y_2^*$) can be calculated.

| | GP | SP | AGE | DENS | PIN | UN | $Y_1$ | $Y_{\cdot,-1}$ | C | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $y_1$ | −0.15<br>(5.6 ) | 0.04<br>(5.2 ) | 0.08<br>(5.7 ) | 0.01<br>(4.4 ) | 0.11<br>(3.4 ) | −0.02<br>(2.5 ) | | 0.81<br>(31.3 ) | −0.21<br>(5.2 ) | 0.888 |
| $y_2$ | −0.17<br>(4.3 ) | 0.10<br>(4.6 ) | 0.24<br>(5.9 ) | 0.18<br>(10.3 ) | 0.04<br>(0.7 ) | −0.21<br>(9.5 ) | 0.02<br>(1.1 ) | 0.25<br>(5.0 ) | 0.21<br>(4.0 ) | 0.860 |

Tab. 1: Estimation results[1])

[1]) T-values in parentheses.

[15]) Note that *DENS* could partly be interpreted as a measure of distance to available specialistic care. Therefore the influence of medical technology is probably less than indicated by the estimated elasticities of *DENS*. A recent "pure" cross-section estimate, based on observations in 83 small regions gives 0.08 as elasticity of $y_1$ with respect to *DENS*.

[16]) The magnitude of the coefficients suggests that more than one factor is responsible for this result. Probably socio-economic differences between the two insured groups play an important role as well.

[17]) In a latter paper we shall analyze this phenomenon in which the differences between the financing systems of the legal and private sector, resp. the capitation system and the fee-for-service system, together with the existing price discrimination, and socio-economic differences between the two insured groups, play an important role.

These long-term elasticities are given in Table 2.

| | GP | SP | AGE | DENS | PIN | UN |
|---|---|---|---|---|---|---|
| $y_1$ | –0.80 | 0.22 | 0.44 | 0.08 | 0.58 | –0.08 |
| $y_2'$[1]) | –1.06 | 0.36 | 0.77 | 0.33 | 0.66 | –0.37 |

[1]) $y_2'$ is the number of continuation cards *per insured.*

Tab. 2: Long-term elasticities

Finally we illustrate the "prediction behavior" of the model in Figure 1, where the predicted and observed development of $y_1$ in the period 1960–1972 for The Netherlands[18]) is given.

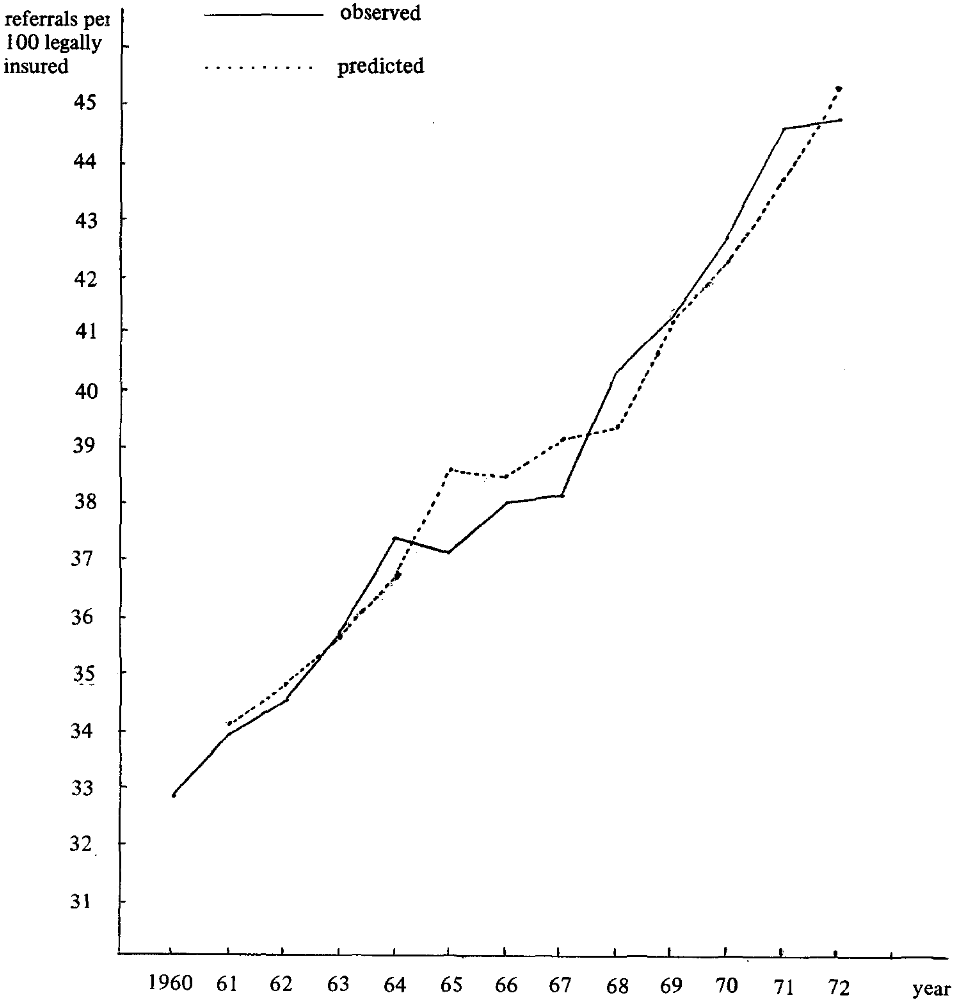


Fig. 1: The predicted and observed behaviour of $y_1$. The "prediction" is based on 1960, given the development of the exogenous variables

## 4. Discussion and Summary

In this paper a dynamic model of aggregated demand for specialistic-outpatient care has been presented. Estimation of the model from a time-series of cross-sections yielded satisfactory well-interpretable results. However, the analysis is far from ideal.
First of all, the measure of "demand" is less precise than desirable, because the number of doctor/patient contacts is only partly reflected by the number of referrals and continuation cards.[19])
Furthermore all specialties are treated together.

Second the equations have to be embodied in a comprehensive model of the Health Care Sector, in which the supply variables will be endogenous. The results of this paper are preliminary results of such a model building project.

More fundamental problems rise with respect to the specification of the model, which depends wholly on "common sense" instead of well-established economic theory.[20]) The functional specification is only chosen for convenience, although it seems to be empirically validated. The arguments in favour of the dynamic specification are plausible, but the restriction to a distributed lag model as chosen can only be seen as a first approach.

Bearing these considerations in mind one must be careful with the interpretation of the estimated coefficients, though they coincide with our expectations and are very significant.

The results can be summarized as follows:

- First, the increase in real consumption of specialistic-outpatient care was partly due to an autonomous increase (results of the changing medical technology) and to the ageing of the population.
- The increase in the supply of specialists yielded a considerable consumption increase as well.
- The only possibility to reduce the increase in the demand for specialistic care is an increase in the number of *GP*'s,[21]) because first level (general) care appears to be a substitute for second level (specialistic) care.
- The dichotomy in the Dutch insurance system seems to have a significant influence on the demand for specialistic care, which for a part may be explained by the different remuneration systems for the publicly and privately insured patients.
  More research on the influence of insurance and remuneration systems on the demand for medical care is needed[22]) in order to guard against these unintended effects.

[18]) i.e., not disaggregated per province; the constant term is corrected in order to see that the disturbances sum to zero.

[19]) Let alone that we know something about the time of a consult, the availability of asistants, the use of technical equipment etc.

[20]) However, "an ounce of practice is generally worth more than a ton of theory", e.g. *Schumacher* [1975, p. 31].

[21]) On first sight it is not clear how in this way *total* medical costs will decrease. However, see *van der Gaag* et al. [1975].

[22]) See for instance the "Health Insurance Project", described in *Newhouse* [1974].

## References

*Arrow, K.J.*: Uncertainty and the Welfare Economics of Medical Care. American Economic Review 53, 1963.

*Balestra, P.*, and *M. Nerlove*: Pooling cross-section and time-series data in the estimation of a dynamic model: the demand for natural gas. Econometrica 34, 1966.

*Brown, J.M.*: Habit Persistence and Lags in Consumer Behaviour. Econometrica 20, 1952.

*Dhrymes, P.J.*: Distributed Lags, Problems of Estimation and Formulation. San Francisco 1971.

*Feldstein, M.S.*: Econometric Studies of Health Economics. Frontiers of Quantitative Economics, II, ed. by M. Intriligator and D. Kendrick. Amsterdam 1974.

*Fuchs, V.*, and *M. Kramer*: Expenditures for Physician's Service in the United States. National Bureau of Economic Research. Occasional paper 117, 1972.

*Gaag, J.v.d., F.F.H. Rutten*, and *B.M.S. Van Praag*: Determinants of hospital utilization in The Netherlands. Health Service Research, Fall, 1975.

*Grether, D.M.*, and *G.S. Maddala*: On the asymptotic properties of some two-step procedures for estimating distributed lag models. International Economic Review 3, 1972.

*Griliches, Z.*: Distributed Lags, a Survey. Econometrica 35, 1967.

*Hatanaka, M.*: An efficient two-step estimator for the dynamic adjustment model with auto-regressive errors. Journal of Econometrics 2, 1974.

*Kmenta, J.*: Elements of econometrics. New York 1971.

*Nerlove, M.*: Further evidence on the estimation of dynamic economic relations from a time-series of cross-sections. Econometrica 39, 1971.

*Newhouse, J.P.*: The Health Insurance Study, a Summary. Rand Corporation R–965–1–OEO, 1974.

*Parks, R.W.*: Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. Journal of American Statistical Association 62, 1967.

*Roemer, M.I.*: On paying the doctor and the implications of different methods. Journal of Health and Social Behaviour, Spring, 1962.

*Sims, A.*: Distributed lags. Frontiers of Quantitative Economics, II, ed. by M.D. Intriligator and D.A. Kendrick, Amsterdam 1974.

*Schumacher, E.F.*: Small is beautiful. London 1975.

*Taylor, L.D.*, and *T.W. Wilson*: Three-pass least squares; a method for estimating models with a lagged dependent variable. The Review of Economics and Statistics 46, 1964.

*Wallis, K.F.*: Lagged dependent variables and serially correlated errors: a reappraisal of three-pass least squares. Review of Economics and Statistics 69, 1967.

*Zellner, A.*: An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association 57, 1962.

Received January 1, 1977
(revised version September 8, 1977)