Forecasting based on Very Small Samples and Additional Non-Sample Information*

Kurt Brännäs and Jörgen Hellström
Department of Economics, Umeå University, Sweden

Umeå Economic Studies 472 (1998)

Abstract

Generalized method of moments estimation and forecasting is introduced for very small samples when additional non-sample information is available. Small simulation experiments are conducted for the linear model with errors-in-variables and for a Poisson regression model. Two empirical illustrations are included. One is based on Ukrainian imports and the other on private schools in a Swedish county.

Key Words: Generalized method of moments, additional information, forecasting, Ukrainian imports, private schools.

JEL classification: C22, C53, C82, F17, I2.

1. Introduction

This paper deals with forecasting in situations when there are only few available observations for the estimation of an underlying econometric model. Such situations are, for instance, encountered when forecasting aspects of developing or transition economies or when forecasting at regional levels for more developed economies. Rather than avoiding the use of econometric models for forecasting we wish to improve on the forecasting performance of such models by capitalizing on additional information in the estimation phase.

The paper introduces generalized method of moments (GMM, Hansen, 1982) estimation for linear and nonlinear models when the sample size is small and additional non-sample information about the model parameters is likely to be most beneficial. The additional or non-sample information about the parameters is taken to be random and is introduced as random linear constraints in the spirit of mixed estimation (ME, Durbin, 1953; Theil and Goldberger, 1961; and others). The additional information can, e.g.,

*The authors gratefully acknowledge constructive comments from seminar participants at Umeå University, University of Lisbon, International Symposium on Forecasting 1998 and The Swedish Econometrics Conference 1998 on an earlier version of the paper.

be purely judgemental or obtained by estimating analogous models for other countries, regions or whatever is regarded appropriate. There are other ways of introducing extra information about parameters, e.g., by inequality restrictions or by adopting a Bayesian approach. Such alternative specifications are not considered, but could be developed along the lines of, e.g., Gourieroux and Monfort (1995, ch. 21). The structural model relating the endogenous variable to exogenous ones can in addition to being possibly nonlinear be either static or dynamic.

The general model framework is set in Section 2, while the GMM estimator and the GMM based forecast are introduced in Section 3. Since the linear static model enables a more detailed analytical study, the research on this particular model is briefly reviewed in Section 4. The Poisson regression model provides a slightly more complex illustration. While analytical studies are feasible for some restricted models, such as the linear, they are not feasible for nonlinear models and for models containing, say, errors-in-variables. Section 4 provides Monte Carlo studies of the estimator and forecasting performance for the linear regression model subject to errors-in-variables and for the nonlinear Poisson regression model. Section 5 contributes by two illustrations based on imports to the Ukraine and on private schools in a Swedish county. A few concluding remarks are saved for the final section.

2. Model

Consider the model

$$y_t = g(\mathbf{z}_t, \boldsymbol{\beta}) + \epsilon_t, \tag{1}$$

where $\mathbf{z}_t = (y_{t-1}, \ldots, y_{t-p}, \mathbf{x}_t, \ldots, \mathbf{x}_{t-q})$, $\boldsymbol{\beta} \in \mathcal{B} \subset \mathcal{R}^k$, and $t = 1, \ldots, T$. The mean function $g(.,.)$ may be linear or nonlinear. The random error term, ϵ_t, is assumed to have zero mean and the covariance matrix of $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_T)'$ is $\boldsymbol{\Sigma}$. Note that a full distributional assumption is not made. This model could without substantial difficulty be generalized to a multivariate one.

In a linear and static model to be estimated by ordinary least squares (OLS), we say that the sample is undersized when $T < k$. When this holds true, the Hessian matrix is not invertible. For nonlinear models noninvertibility of the Hessian matrix suggests that the sample is undersized and/or that lack of identification due to an unfortunate parametrization are indicated.

The additional information is provided in the form[1]

$$\mathbf{q} = \mathbf{R}\boldsymbol{\beta} + \boldsymbol{\zeta}, \tag{2}$$

where $\mathbf{q}$ is an observed $(m \times 1)$ vector, the $(m \times k)$ matrix $\mathbf{R}$ is given, and $\boldsymbol{\zeta}$ is an unobserved random error with zero mean vector and a given covariance matrix $\boldsymbol{\Omega}$. The m is the number of additional information sources. In addition, we make the assumption that $E(\epsilon_t \boldsymbol{\zeta}) = \mathbf{0}$.

[1]See, e.g., Kennedy (1991) for a recent and interesting example.

3. Estimation and Forecasting

The estimation approach is GMM. The feature that makes the present setup different from the conventional one based on (1), is that there are two sets of moment conditions; those based on (1) and those based on (2).

For the model we consider moment conditions of the form

$$\mathbf{m}_1 = T^{-1}\mathbf{U}'(\mathbf{y} - g(\mathbf{Z}, \boldsymbol{\beta})). \tag{3}$$

Here, $g(\mathbf{Z}, \boldsymbol{\beta})$ has row elements $g(\mathbf{z}_t, \boldsymbol{\beta})$ and $\mathbf{U}$ is a $(T \times n)$ matrix of instrumental variables, with $n \geq k$. The content of $\mathbf{U}$ should depend on the specification of $g(.,.)$. For instance, in a static model (with $\mathbf{z}_t = \mathbf{x}_t$) the $\mathbf{x}_t$ may serve as its own instrument, while if $\mathbf{x}_t$ is endogenous or measured with error some other instrumental variable is required (see also Lewbel, 1997). In the classical setup, plim $\mathbf{m}_1 = E(\mathbf{m}_1) = \mathbf{0}$ is used to justify the GMM estimator.

For the additional information about the parameter vector, the moment condition is written

$$\mathbf{m}_2 = m^{-1}\mathbf{V}'(\mathbf{q} - \mathbf{R}\boldsymbol{\beta}), \tag{4}$$

where $\mathbf{V}$ is an $(m \times p)$ matrix of instrumental variables. In this instance, we norm by the number, m, of replicates of additional information sources.

Since $\mathbf{m}_1$ and $\mathbf{m}_2$ are independent by the assumption $E(\epsilon_t \boldsymbol{\zeta}) = \mathbf{0}$, we may write the GMM criterion on an additive form, yielding

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \arg\min_{\boldsymbol{\beta} \in \mathcal{B}} \{\mathbf{m}_1'\mathbf{W}_1^{-1}\mathbf{m}_1 + \mathbf{m}_2'\mathbf{W}_2^{-1}\mathbf{m}_2\} \\ &= \arg\min_{\boldsymbol{\beta} \in \mathcal{B}} \{\boldsymbol{\epsilon}'\mathbf{U}(\mathbf{U}'\boldsymbol{\Sigma}\mathbf{U})^{-1}\mathbf{U}'\boldsymbol{\epsilon} + \boldsymbol{\zeta}'\mathbf{V}(\mathbf{V}'\boldsymbol{\Omega}\mathbf{V})^{-1}\mathbf{V}'\boldsymbol{\zeta}\}. \end{aligned}$$

The asymptotic covariance matrix of the estimator is then of the form

$$Cov(\hat{\boldsymbol{\beta}}) = (\mathbf{D}'\mathbf{U}(\mathbf{U}'\boldsymbol{\Sigma}\mathbf{U})^{-1}\mathbf{U}'\mathbf{D})^{-1} + (\mathbf{R}'\mathbf{V}(\mathbf{V}'\boldsymbol{\Omega}\mathbf{V})^{-1}\mathbf{V}'\mathbf{R})^{-1},$$

where $\mathbf{D} = -\partial g(\mathbf{Z}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}'$ is evaluated at estimates.

Since the case of interest is when the sample size is very small, large sample properties of the estimator, such as consistency, can only be taken as rough approximations. On the other hand, small sample properties are difficult to study except for in some special cases. Small-σ (e.g., Kadane, 1971) and related approaches to find approximate properties has not been found useful at this stage. For practical implementation a first stage based on, e.g., the artificial settings $\boldsymbol{\Sigma} = \mathbf{I}$ and $\boldsymbol{\Omega} = \mathbf{I}$ can be used, or $\boldsymbol{\Sigma}$ can be estimated directly by excluding the additional information part. With $\boldsymbol{\beta}$ estimated an estimate of $\boldsymbol{\Sigma}$ can be obtained from the model part and then used together with the given $\boldsymbol{\Omega}$ in a second step. Below additional light is cast on estimator performance for specialized models.

The h-steps-ahead forecast

$$\hat{y}_{T+h|T} = E[g(\mathbf{z}_{T+h}^{\mathbf{0}}, \boldsymbol{\beta})|Y_T]$$

is evaluated at estimates $\hat{\boldsymbol{\beta}}$, $Y_T = (y_1, \ldots, y_T)$ is the information set, and $\mathbf{z}^0_{T+h} = (y_{T+h-1}, \ldots, y_{T+h-p}, \mathbf{x}^0_{T+h}, \ldots, \mathbf{x}^0_{T+h-q})$ with $\mathbf{x}^0_{T+j}, j = 1, \ldots, h$, is assumed known. For nonlinear dynamic models the evaluation of the conditional expectation may be a very difficult problem (e.g., Granger and Teräsvirta, 1993, ch. 8).

Using a first order Taylor expansion about the true parameter vector $\boldsymbol{\beta}_0$, the forecast error is

$$\begin{aligned}
\hat{e}_{T+h} &= y_{T+h} - \hat{y}_{T+h|T} \\
&\approx g(z^0_{T+h}, \boldsymbol{\beta}_0) - E[g(\mathbf{z}^0_{T+h}, \boldsymbol{\beta}_0)|Y_T] - \frac{\partial E[g(\mathbf{z}^0_{T+h}, \boldsymbol{\beta}_0)|Y_T]}{\partial \boldsymbol{\beta}'}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \epsilon_{T+h}.
\end{aligned}$$

From the approximate forecast error we have the general properties:

$$\begin{aligned}
E(\hat{e}_{T+h}) &= E_{Y_T}[E(\hat{e}_{T+h}|Y_T)] = -E_{Y_T}\left[\frac{\partial E[g(\mathbf{z}^0_{T+h}, \boldsymbol{\beta}_0)|Y_T]}{\partial \boldsymbol{\beta}'}(E(\hat{\boldsymbol{\beta}}|Y_T) - \boldsymbol{\beta}_0)\right] \\
V(\hat{e}_{T+h}) &= V_{Y_T}[E(\hat{e}_{T+h}|Y_T)] + E_{Y_T}[V(\hat{e}_{T+h}|Y_T)] \\
&= V_{Y_T}\left[\frac{\partial E[g(\mathbf{z}^0_{T+h}, \boldsymbol{\beta}_0)|Y_T]}{\partial \boldsymbol{\beta}'}E(\hat{\boldsymbol{\beta}}|Y_T)\right] + \sigma^2 + E_{Y_T}[V(g(z^0_{T+h}, \boldsymbol{\beta}_0)|Y_T)] \\
&\quad + E_{Y_T}\left[\frac{\partial E[g(\mathbf{z}^0_{T+h}, \boldsymbol{\beta}_0)|Y_T]}{\partial \boldsymbol{\beta}'}Cov(\hat{\boldsymbol{\beta}}|Y_T)\frac{\partial E[g(\mathbf{z}^0_{T+h}, \boldsymbol{\beta}_0)|Y_T]}{\partial \boldsymbol{\beta}}\right],
\end{aligned}$$

where $\sigma^2 = V(\epsilon_{T+h})$.

For particular models these expressions may be simplified and then easier to interpret. For instance, when there are no lagged endogeneous variables and the exogenous variables are fixed, i.e. when $\mathbf{z}^0_{T+h} = (\mathbf{x}^0_{T+h}, \ldots, \mathbf{x}^0_{T+h-q})$, we obtain the predictor $\hat{y}_{T+h|T} = g(\mathbf{z}^0_{T+h}, \boldsymbol{\beta})$ and

$$\begin{aligned}
E(\hat{e}_{T+h}) &= -\frac{\partial g(\mathbf{z}^0_{T+h}, \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}'}(E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}_0) \\
V(\hat{e}_{T+h}) &= \sigma^2 + \frac{\partial g(\mathbf{z}^0_{T+h}, \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}'}Cov(\hat{\boldsymbol{\beta}})\frac{\partial g(\mathbf{z}^0_{T+h}, \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}.
\end{aligned}$$

If in addition the model is linear in parameters, $\partial g(\mathbf{z}^0_{T+h}, \boldsymbol{\beta}_0)/\partial \boldsymbol{\beta}' = \mathbf{z}^0_{T+h}$.

4. Two Special Models

4.1 The Linear Regression Model

Durbin (1953), Theil (1963) and others consider the linear model by augmenting the data set by artificial observations corresponding to the random additional information about parameters.

Suppose we wish to estimate the parameters of the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{y}$ is the T vector of observations on the endogeneous variable, $\mathbf{X}$ is the matrix of fixed exogenous variables, $\boldsymbol{\beta}$ is the parameter vector, and $\boldsymbol{\epsilon}$ is the vector of disturbances with mean zero and known covariance matrix $\mathbf{\Sigma}$. The additional information on $\boldsymbol{\beta}$ is available in the form of random linear constraints: $\mathbf{q} = \mathbf{R}\boldsymbol{\beta} + \boldsymbol{\zeta}$.

Combining the two independent sources of sample and extraneous information we have

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{R} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\zeta} \end{pmatrix}.$$

The covariance matrix of the disturbance term is block diagonal with blocks $\mathbf{\Sigma}$ and $\mathbf{\Omega}$, respectively. The generalized least squares (GLS), the mixed and the GMM (with $\mathbf{U} = \mathbf{X}$ and $\mathbf{V} = \mathbf{I}$) estimators are all equal, such that

$$\hat{\boldsymbol{\beta}}_{ME} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X} + \mathbf{R}'\mathbf{\Omega}^{-1}\mathbf{R})^{-1}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y} + \mathbf{R}'\mathbf{\Omega}^{-1}\mathbf{q}).$$

This estimator is best linear and unbiased with covariance matrix

$$V(\hat{\boldsymbol{\beta}}_{ME}) = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X} + \mathbf{R}'\mathbf{\Omega}^{-1}\mathbf{R})^{-1}.$$

Note that in the absence of data on $\mathbf{y}$ and $\mathbf{X}$ these expressions reduce to those of a GLS estimator based on only the additional information, while without additional information the conventional GLS estimator emerges. As the variance of the random additional information goes to zero, the mixed estimator approaches the restricted LS estimator (e.g., Fomby et al., 1984, ch. 6) implied by the exact restriction $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$. On the other hand, as the random restrictions become less certain and assuming that $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2\mathbf{I}$, the mixed estimator approaches the ordinary least squares (OLS) estimator. With $\mathbf{\Omega}$ known through the external source, restricted versions of $\mathbf{\Sigma}$ can be estimated by some two stage procedure. Theil (1963) offers a large sample justification, and Swamy and Mehta (1969) study the finite sample properties.

The predictor of y_{T+h} using mixed estimation and based on known $\mathbf{x}^0_{T+h}$ is $\hat{y}^0_{ME} = \mathbf{x}^0_{T+h}\hat{\boldsymbol{\beta}}_{ME}$. The forecast error is $e^0_{ME} = \mathbf{x}^0_{T+h}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{ME}) + \epsilon_{T+h}$ so that $E(e^0_{ME}) = 0$, and the forecast error variance is

$$V(e^0_{ME}) = \sigma^2 + \mathbf{x}^0_{T+h}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X} + \mathbf{R}'\mathbf{\Omega}^{-1}\mathbf{R})^{-1}\mathbf{x}^{0'}_{T+h}.$$

When the basic assumptions are satisfied and the additional information is unbiased then forecasting based on ME offers a gain.

If $\mathbf{X}$ or some part of $\mathbf{X}$ is measured with error the mixed estimator will be biased. Measurement errors appears a real problem for transition and developing economies, with new definitions and measurement practices being put into place. An instrumental variable (IV) or GMM estimator for the linear model is easily obtained but its properties in small samples are largely unknown. The IV estimator and its covariance matrix are of the form

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{IV} &= (\mathbf{X}'\mathbf{A}\mathbf{X} + \mathbf{R}'\mathbf{\Omega}^{-1}\mathbf{R})^{-1}(\mathbf{X}'\mathbf{A}\mathbf{y} + \mathbf{R}'\mathbf{\Omega}^{-1}\mathbf{q}) \\ Cov(\hat{\boldsymbol{\beta}}_{IV}) &= (\mathbf{X}'\mathbf{A}\mathbf{X} + \mathbf{R}'\mathbf{\Omega}^{-1}\mathbf{R})^{-1} \end{aligned}$$

with $\mathbf{A} = \mathbf{U}(\mathbf{U}'\mathbf{\Sigma}\mathbf{U})^{-1}\mathbf{U}'$ and where $\mathbf{U}$ is the instrumental variable. For the IV estimator there are equal numbers of variables in $\mathbf{U}$ and $\mathbf{X}$, i.e. $n = k$. The GMM estimator is more general in the sense that it can encompass more variables in $\mathbf{U}$ than in $\mathbf{X}$, i.e. $n \geq k$. Lewbel (1997) shows how to construct instrumental variables from only the variables contained in the model to attain consistency of the estimator. Obviously, this is a potentially very useful approach to estimation, whenever it is difficult to find external instrumental variables. There is some doubt, however, that the estimator performs well in small samples.

4.1.1 Simulation Experiment

To cast some light on the small sample properties of the GMM estimator for a case of measurement errors in $\mathbf{X}$, and to study coverage probabilities of forecast confidence intervals we perform a small simulation experiment. The assumed models for the generation of data are of the form

$$y_t = \beta_0 + \beta_1 x_t^* + \epsilon_t \quad \text{and} \quad \mathbf{q} = \boldsymbol{\beta} + \boldsymbol{\zeta}.$$

The parameters are set at $\beta_0 = 1$ and $\beta_1 = 0.2$. The unknown x_t^* is generated from an AR(1) model ($x_t^* = 0.7x_{t-1}^* + \upsilon_t$) and level shifted by adding 5 to x_t^*. The measurement error is introduced as $x_t = x_t^* + \eta_t$ and the instrumental variable is generated as $u_t = 0.7x_t^* + \xi_t$. Additional information about $\boldsymbol{\beta}$ is available with $m = 2, 4, 6, 8, 10$ and 12. The random errors υ_t, η_t and ξ_t are generated independently from $N(0, 1)$ distributions. The $\zeta_i, i = 1, 2$, are generated as $N(0, \sigma_\zeta^2)$ with $\sigma_\zeta^2 = 0.05$ and 0.1, and ϵ_t as $N(0, 2)$, and mutually independent as well as independent of other error terms. Sample size is varied; $T = 3, 6, 9, 12, 15$ and 18, and 1000 replications are run in each cell. Note that the explanatory power in both models is low. For $\sigma_\zeta^2 = 0.1$ the range of the $\beta_i, i = 0, 1$, is approximately 1.26 (using $\sigma_\zeta \approx$ Range/4), while for $\sigma_\zeta^2 = 0.05$ the range is approximately 0.89. For the y_t model a low theoretical R^2 may give a higher estimated R^2 for few rather than for many observations.

Besides estimation by GMM (or equivalently by IV) using the u_t as an instrument, the OLS estimator based on $y_t = \beta_0 + \beta_1 x_t + (\epsilon_t - \beta_1 \eta_t)$ is applied. The OLS estimator is also used to obtain estimates of σ^2 for GMM estimation.

Some indicative bias and MSE results for β_1 are given in Figure 1. As expected the MSEs of both estimators get smaller as T increases. For the GMM estimator the bias is larger for the larger sample size and small m. An explanation to this lies in the rather low explanatory power of the model. As T increases the less precise model observations become relatively more important. Both measures are decreasing in the number of additional information 'observations', m. In terms of bias, OLS is doing better than GMM when this utilizes additional information that is not precise, i.e. when σ_ζ^2 is large. For MSE, the particular extra information used here brings about improvements for all T. The improvement is largest for small T and m. This is the region where improvements are of most value. These conclusions hold for other parameter combinations as well.

Figure 2 reports coverage probabilities for forecast confidence intervals. In calculating the intervals quantiles are obtained from the $t(T + m - 2)-$distribution for the GMM estimator and from the $t(T - 2)-$distribution for the OLS estimator. The coverage probabilities for OLS based forecasts are not significantly different from the nominal 0.95 level.

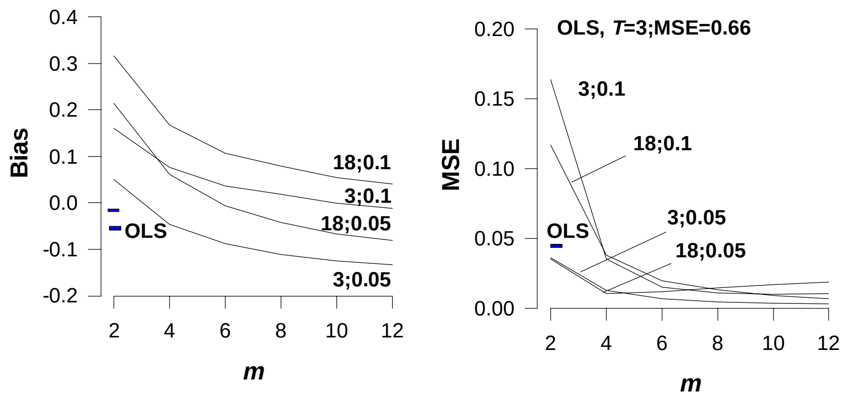

Figure 1: Bias and MSE for β_1 using GMM and OLS estimation at $T;\sigma_\zeta^2$.

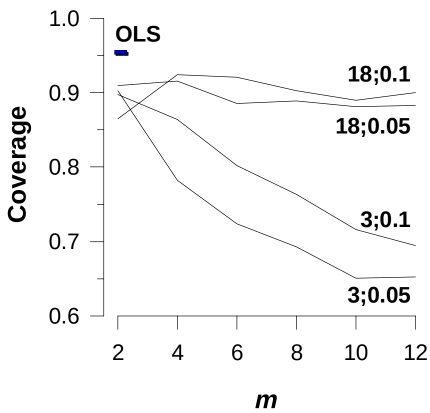

Figure 2: Coverage probabilities for forecast confidence intervals based on GMM and OLS estimation at $T;\sigma_\zeta^2$.

For forecasts based on the GMM estimator the discrepancy from the nominal level is significant. Less precise additional information yields coverage probabilities closer to the nominal level and so does a larger T.

4.2 The Poisson Regression Model

The Poisson regression model has both the mean and the variance of independent y_t equal to $\lambda_t = \exp(\mathbf{x}_t\boldsymbol{\beta})$, where $\mathbf{x}_t$ is a vector of exogenous variables. Hence, in this case the ϵ_t of eq. (1) has variance λ_t, so that heteroskedasticity is an essential feature of the model. A number of different moment restrictions, e.g., those that correspond to the estimators considered by Gourieroux et al. (1984), can be utilized for estimation. We consider restrictions based on the likelihood equation, i.e. $\mathbf{X}'(\mathbf{y}-\boldsymbol{\lambda}) = \mathbf{0}$ with $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$, to obtain the criterion function

$$(\mathbf{y}-\boldsymbol{\lambda})'\mathbf{X}(\mathbf{X}'\boldsymbol{\Lambda}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y}-\boldsymbol{\lambda}) + (\mathbf{q}-\mathbf{R}\boldsymbol{\beta})'\mathbf{V}(\mathbf{V}'\boldsymbol{\Omega}\mathbf{V})^{-1}\mathbf{V}'(\mathbf{q}-\mathbf{R}\boldsymbol{\beta}),$$

where $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$. Since the first order condition is nonlinear in $\boldsymbol{\beta}$ no explicit solution exists. The best forecast can be obtained by specializing a result for serially correlated and overdispersed $y_1, \ldots, y_T$ (Brännäs, 1995a) and is here of the simple form $\boldsymbol{\lambda}_{T+h}$.

Consider as an example the regressor free case so that λ is constant and $\boldsymbol{\Omega} = \omega\mathbf{I}_m$. Then the natural (in the sense that it still yields a simple linear estimator) criterion to minimize is of the form

$$\sum_{t=1}^{T}(y_t-\lambda)^2/\lambda_0 + \sum_{j=1}^{m}(q_j-\lambda)^2/\omega,$$

where λ_0 is fixed at some initial value. Minimization yields the estimator

$$\hat{\lambda} = \left[\lambda_0^{-1}\sum_{t=1}^{T} y_t + \omega^{-1}\sum_{j=1}^{m} q_j\right]/u,$$

where $u = T/\lambda_0 + m/\omega$. The estimator has the properties $E(\hat{\lambda}) = \lambda$ and $V(\hat{\lambda}) = \lambda(T/\lambda_0^2 + m/\omega^2)/u^2$.

Hence, the estimator is unbiased and as $\omega \to \infty$, the estimator has a variance approaching that of the maximum likelihood (ML) estimator. For smaller ω, Figure 3 presents illustrations showing that the MSE of this mixed estimator may be much smaller than that of the ML estimator. For $\omega \to 0$, the MSEs approach λ/m.

The best forecast is $\hat{\lambda}$ and the forecast variance is equal to $\lambda[1 + (T/\lambda_0^2 + m/\omega^2)/u^2]$. Hence, forecast properties depend on the quality of the additional information in much the same way as the estimator.

4.2.1 Simulation Experiment

To illustrate the properties of the estimator and the forecast for a more general Poisson model we conduct a Monte Carlo experiment. The assumed model is Poisson with

$$\lambda_t = \exp(\beta_0 + \beta_1 x_t).$$

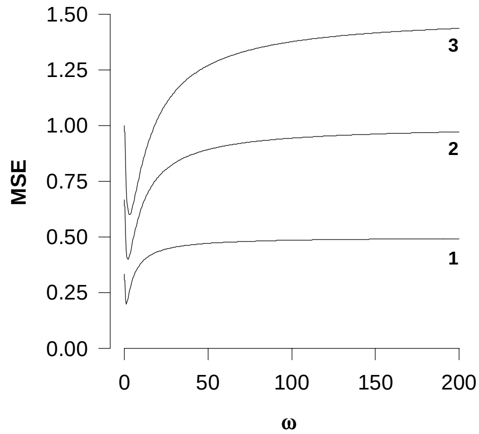

Figure 3: MSEs for Poisson model with $\lambda = \lambda_0 = 1, 2, 3, T = 2, m = 3$ and ω varied.

The parameters are set at $\beta_0 = 1$ and $\beta_1 = 0.2$. The x_t is generated as $N(5, 4)$ but kept fixed over the 1000 replications run in each cell, and the sample size is varied; $T = 3, 6, 9, 12, 15$ and 18. The additional information about $\boldsymbol{\beta}$ is available in the form of $\mathbf{q} = \boldsymbol{\beta} + \boldsymbol{\zeta}$, with $m = 2, 4, 6, 8, 10$ and 12. Here, $\zeta_i, i = 1, 2$, is generated as $N(0, \sigma_\zeta^2)$, with $\sigma_\zeta^2 = 0.5$ and 1. Note that $\mathbf{q}$ generated this way carries very little information.

Some indicative bias and MSE results for β_1 are given in Figure 4. As expected both the bias and the MSE get smaller as T increases. Both measures are constant or increasing in the number of extra information 'observations', m, for larger sample sizes. In terms of bias, ML is best already at $T = 3$ observations. For MSE the particular extra information used here appears to bring about improvements for all T.

Figure 5 reports empirical coverage probabilities for forecast confidence intervals based on the normal distribution. Coverage probabilities get closer to the nominal 0.95 level as T increases, while a larger m appears to have a very small effect. On comparison with the ML based forecasts, coverage probabilities based on GMM estimation are higher for $T \leq 9$, while for larger values there are no substantial differences. Coverage probabilities are significantly too small, with the exception of the ML based interval for $T = 18$. The forecasts appear to be (not significantly so) downward biased and to have skewed distributions for small T. Using intervals based on the $t-$distribution would increase the empirical coverage probabilities.

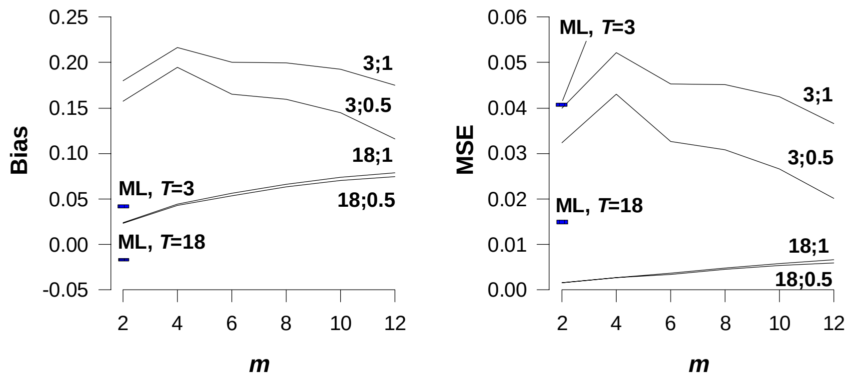

Figure 4: Bias and MSE for β_1 using GMM and ML estimation at $T; \sigma_\zeta^2$.

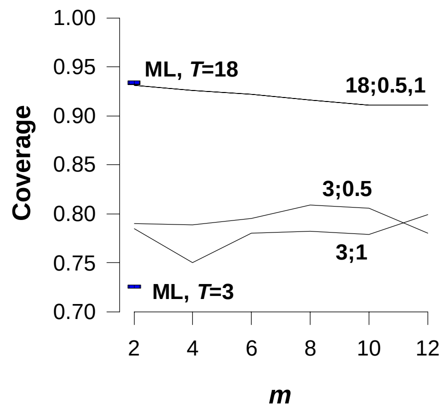

Figure 5: Coverage probabilities for forecast confidence intervals based on GMM and ML estimation at $T; \sigma_\zeta^2$.

5. Empirical Illustrations

5.1 Import Demand

As an illustration we present a model based forecast for imports to the Ukraine. A feature of this problem is a small sample and a potentially low data quality. The small sample size is typical of economic activity in many transitional countries.

The econometric model is based on the economic theory of import demand (e.g., Goldstein and Kahn, 1985). The simplest form of import demand models consistent with theory is

$$M_t = f(P_t, Y_t),$$

where M_t is the quantity of import demanded in period t, P_t is the ratio of price of imports relative to the domestic price level, and Y_t is the real gross domestic product (GDP). For developed countries a negative sign is expected for $\partial M/\partial P$, while $\partial M/\partial Y$ is expected to be positive. For a transition economy as the Ukrainian one we may expect $\partial M/\partial Y$ to be negative. An explanation to this is the growing demand for imported goods in the first transitional stages when the economy is in recession.

A dynamic behaviour could be introduced by including lagged dependent and/or independent variables linearly or log-linearly. Thursby and Thursby (1984) conclude that the model should include some dynamic behaviour, preferably a lagged dependent variable. Boylan et al. (1979) conclude that the log-linear form is to be preferred. Due to the small sample size, we specify a static model in logarithmic variables.

With respect to the choice between nominal or real variables we choose the former (cf. Branson, 1968). By this we avoid having to explain two series, one for the quantity of imports and one for the import prices in order to get the current value. The chosen approach has some drawbacks. The determinants of prices and volume are different and a single equation carries the danger that the estimated coefficients will contain some interaction of supply and demand influences.

The model, in a logarithmic functional form, is specified as

$$\log M_t = \beta_0 + \beta_1 \log Y_t + \beta_2 \log P_t + \epsilon_t,$$

where M_t and Y_t are aggregated nominal values of imports and GDP, respectively. The data is mainly collected from International Financial Statistics (IMF, 1997) and covers the years 1992–1995. Since an import price index did not exist for the Ukraine, one was constructed from foreign export prices and the share of export to the Ukraine. All data are transformed into USD. The activity variable is probably underestimated, since transitional countries often experience rapid growth of the informal sector during the first stage of transition.

The model is first estimated by OLS on the four annual observations for the Ukraine. These estimates and others are given in Table 1. The estimated effect of GDP is negative and significant, while the price variable has an expected and significant effect. The prior information about the parameters of the model are obtained by estimating corresponding models for Poland and Hungary on the years 1990–1994. Combinations of the prior information are also reported in Table 1. Using both countries the mixed estimator effects of

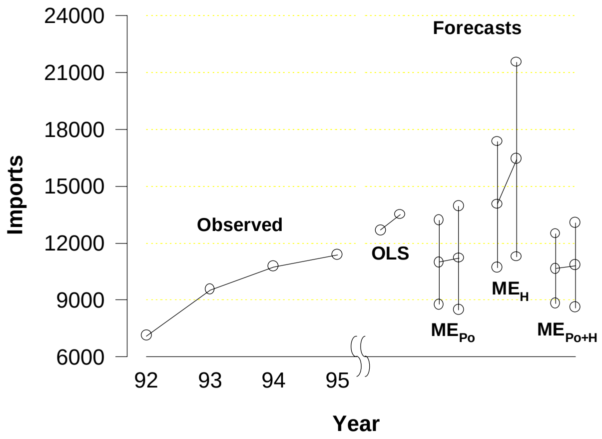

Figure 6: Observed Ukrainian imports and forecasts with 95 percent forecast intervals.

Y and P are -0.27 and -0.23, respectively. The standard errors of the mixed estimators are smaller. An attempt to estimate the parameters by a fixed effect panel data approach gives quite unreasonable estimates.

To produce ex ante forecasts independent variables are first forecasted two years ahead using an AR(1) model. The resulting forecasts and 95 percent confidence intervals (based on the t–distribution) are presented in Figure 6. Note that the confidence interval for the pure OLS estimator is based on the $t(1)$–distribution. This implies a length of the OLS intervals that stretches beyond the figure.

Table 1: Estimation results for import models (standard errors in parentheses).

	Ukraine				Poland	Hungary
Variable	OLS	$\mathrm{ME_{Po}}$	$\mathrm{ME_H}$	$\mathrm{ME_{Po+H}}$	$\mathrm{OLS_{Po}}$	$\mathrm{OLS_H}$
$\log Y$	-0.45	-0.35	0.11	-0.27	-0.27	0.34
	(0.17)	(0.05)	(0.12)	(0.05)	(0.05)	(1.22)
$\log P$	-0.43	-0.30	-0.11	-0.23	0.06	0.38
	(0.09)	(0.04)	(0.07)	(0.04)	(0.07)	(0.31)
Constant	14.96	13.51	8.50	12.56	12.17	3.84
	(1.88)	(0.19)	(1.40)	(0.18)	(0.88)	(11.49)
s^2	0.0022	0.0116	0.0147	0.0130	0.0829	0.0117

5.2 Private Schools

Estimates and forecasts for the entry and exit of elementary private schools are given for the county of Västerbotten, Sweden. The sample consists of only 45 observations on 15 municipalities with three annual observations on the number of public schools. Overall there are only a few non-zero observations on the dependent variable. A school finance reform (1991-1992) has stimulated the entry of private schools in Sweden. From 1991/1992 up till now the total number for the country has increased from 90 till 350. While the debate on the role of private schools as providers of education is not new, relatively little is known about the economic incentives or disincentives for private schools. Hoxby (1993) found evidence that public schools improve their quality in communities where competition from private schools is strenuous. Downes and Greenstein (1996) shed some light on the determinants of localization choice of private schools in California during 1978-1979.

We model the number of private schools by an integer-valued autoregressive model of order one (e.g., Brännäs, 1995b). This model may, e.g., be written on the form

$$y_{it} = \pi_i y_{it-1} + \phi_i + \epsilon_{it},$$

where y_{it} denotes the number of private schools in the ith municipality at time t, $\pi = 1/(1+\exp(\theta))$ denotes the survival probability of the existing private schools,[2] ϕ_i is the mean entry and the error term ϵ_{it} has zero mean. The mean entry ϕ_i could be modelled, e.g., as the mean function of a Poisson variable ($\xi_{it} = \phi_i + \epsilon_{it}$). In this particular case there appears to be little room for elaborate behavioural models, and we let $\phi_i = \phi + \alpha$ for the largest municipality (Umeå) of the county and let $\phi_i = \phi$ for all other municipalities. In general, ϕ_i could be a function of public school characteristics and the regional demographic structure (Downes and Greenstein, 1996, and references therein).

The sample is obtained from the National Board of Education (Skolverket) and covers the calendar years 1992-1994 for 286 Swedish municipalities. Since the model is dynamic there will be only two observations for the two parameters for each municipality. Moreover there is no variation within most of the municipalities while there is more variation across municipalities. The model for Västerbotten was first estimated by nonlinear least squares (NLSQ). Second, extra information was obtained for the θ and ϕ parameters by estimating the model without a dummy variable (cf. α) on other Swedish municipalities and used in a final step for GMM estimation. The estimation results are presented in Table 2 and indicate substantial improvement in estimation efficiency and that the survival probability is very close to one. The mean entry $\hat{\phi}$ is slightly higher when the extra information is used, while $\hat{\alpha}$ is slightly smaller.

Since $\hat{\pi} \approx 1.0$ it follows (Brännäs, 1995b) that the best forecast for the number of public schools is the final observation, i.e. $y_{i,1995}$. As $\hat{\phi} = 0.08$ we forecast an entry about every 12 years for the other municipalities. Using $\hat{\phi}$ and $\hat{\phi} + \hat{\alpha}$ as well as their confidence intervals we may obtain the corresponding density function evaluated at $\hat{\phi}$ and $\hat{\phi} + \hat{\alpha}$ as well as at the lower and upper limits of their confidence intervals. We find that for no entries in the next year the probability is between 0.20 and 0.24 for Umeå while between

[2] $1 - \pi$ is the exit probability. Note that π is kept constant across municipalities.

Table 2: Estimation results for private school models (standard errors in parentheses).

Para-meter	Västerbotten NLSQ	Västerbotten GMM	Sweden NLSQ
ϕ	0.07	0.08	0.20
	(0.13)	(0.01)	(0.10)
α	1.43	1.42	–
	(0.52)	(0.05)	
θ	-10.86	-20.30	-14.63
	(301.84)	(461.2)	(58.74)

0.91 and 0.94 for the other municipalities. From these values we easily find probabilities for one or more new private schools. Once a new private school is founded it will according to the model survive.

6. Conclusions

The paper has demonstrated that additional information may be incorporated for the estimation of more general models than the previous limitation to linear models have indicated. The additional information may come from different sources. When data is available for corresponding phenomena for other countries, regions, etc. estimation results for these may be utilized using the present framework. An alternative is then obviously to employ a panel data approach instead. The additional information may also take the form of subjective judgemental assessments, which then breaks the ties to panel data estimation.

The Monte Carlo simulations indicate that the use of additional information is most beneficial when the sample size is at its smallest. A realistic goal for forecasting performance in circumstances that we have tried to face in this paper should obviously be placed lower than for cases with long and stable time series data. Using models and clear-cut estimation procedures, we believe, makes it easier to communicate, evaluate and improve on forecasting practise.

References

Boylan, T.A., Cuddy, M.P. and O'Muircheartaigh, J. (1980). The Functional Form of the Aggregate Import Demand Equation: A Comparison of Three European Economies. *Journal of International Economics* **10**, 561-566.

Brännäs, K. (1995a). Prediction and Control for a Time Series Count Data Model. *International Journal of Forecasting* **11**, 263-270.

Brännäs, K. (1995b). Explanatory Variables in the AR(1) Poisson Model. Umeå Economic Studies 381 (revised).

Branson, W.H. (1968). A Disaggregated Model of the U.S. Balance of Trade. Staff Economic Studies No. 44, Board of Governors of the Federal Reserve System.

Downes, A.T. and Greenstein, M.S. (1996). Understanding the Supply Decisions of Nonprofits: Modelling the Location of Private Schools. *RAND Journal of Economics* **27**, 365-390.

Durbin, J. (1953). A Note on Regression when there is Extraneous Information about one of the Coefficients. *Journal of the American Statistical Association* **48**, 799-808.

Fomby, T.B., Hill, R.C. and Johnson, S.R. (1984). *Advanced Econometric Methods.* Springer-Verlag, New York.

Goldstein, M. and Khan, M.S. (1985). Income and Price Effects in Foreign Trade, *Handbook of International Economics*, *Vol 2.* North-Holland, Amsterdam.

Gourieroux, C. and Monfort, A. (1995). *Statistics and Econometric Models.* Cambridge University Press, Cambridge.

Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Applications to Poisson Models. *Econometrica* **52,** 701-720.

Granger, C.W.J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships.* Oxford University Press, Oxford.

Hansen, L.P. (1982), Large Sample Properties of Generalized Method of Moment Estimators. *Econometrica* **50,** 1029-1054.

Hoxby, C.M. (1993). Do Private Schools Provide Competition for Public Schools? Mimeo, Massachusetts Institute of Technology.

International Monetary Fund. (1997). *Yearbook.*

Kadane, J.B. (1971). Comparison of k-class Estimators when the Disturbances are Small. *Econometrica* **39**, 723-737.

Kahn, M.S. and Ross, K.Z. (1977), The Functional Form of Aggregate Import Demand. *Journal of International Economics* **7**, 149-160.

Kennedy, P. (1991). An Extension of Mixed Estimation, with an Application to Forecasting New Product Growth. *Empirical Economics* **16**, 401-415.

Lewbel, A. (1997). Constructing Instruments for Regressions with Measurement Error when no Additional Data are Available, with an Application to Patents and R&D. *Econometrica* **65**, 1201-1213.

Swamy, P.A.V.B. and Mehta, J.S. (1969). On Theil's Mixed Regression Estimator. *Journal of the American Statistical Association* **64**, 273-276.

Theil, H. (1963). On the Use of Incomplete Prior Information in Regression Analysis. *Journal of the American Statistical Association* **58**, 401-414.

Theil, H. and Goldberger, A.S. (1961). On Pure and Mixed Statistical Estimation in Economics. *International Economic Review* **2**, 65-78.

Thursby, J. and Thursby, M. (1984). How Reliable are Simple, Single Equation Specifications of Import Demand? *Review of Economics and Statistics* **66**, 120-128.