Bounded-Influence Estimation Techniques for the Analysis of Structural Macroeconometric Models

Marilena Furno and Christopher F. Baum
University of Naples   Boston College

Working Paper no. 163
March, 1988

0. INTRODUCTION AND OVERVIEW

Considerable attention has been focused in recent years on the difficulties of structural econometric modelling in the presence of structural change, parameter instability (e.g. the Lucas critique) and autoregressive heteroskedasticity. We apply new developments in the field of robust estimation — specifically, the bounded–influence instrumental variables estimator (BIV) — as a diagnostic tool to identify the presence of these weaknesses in the structural estimates of macroeconometric models.

Robust techniques have been put forth as alternatives to the standard least squares procedures in the presence of various departures from the classical assumptions underlying least squares techniques (and, by extension, much of the work that is and has been done with structural econometric modelling of the macroeconomy). We see their potential not as alternative methods, but as more sophisticated diagnostic tools which may be put to use to readily identify these departures, and lead to their correction. Application of the BIV estimators produces several useful byproducts which can lead the researcher to the identification of, for instance, structural change in the model, or a drift in the parameters of the error distribution. These diagnoses may then be used to produce more reliable estimates via standard techniques.

Examples are given from Klein's Model I and Fair's small linearisation of the Fair U.S. Model which illustrate how these techniques might profitably be applied. Our implementation does not require specialised software, as it is implemented in the well–known RATS econometric package for mainframe and micro–computers.

The plan of the paper is as follows. Section 1 presents a short summary of the rationale for robust estimation techniques as an adjunct to classical least squares methods, with particular emphasis on their applicability to simultaneous equations models encountered in applied econometric research. The key elements of the Bounded–Influence Instrumental Variables estimator, due to Krasker and Welsch (1985), are presented, and the extension of analytical results for that estimator by Furno (1988a) are summarised. In Section 2, a money demand equation from Fair's small U.S. model (1984) is presented as an example of structural change in a single–equation context; robust techniques are utilised to demonstrate that the equation still lacks desirable properties following the introduction of a shift dummy. The third and fourth sections present applications of the BIV estimator to two small econometric models – Klein's Model I (1950) and Fair's small model – and considers the performance of equations estimated via BIV techniques in the context of simulation and forecasting. The last section summarises the paper, and presents suggestions for further research.

1. ROBUST ESTIMATION TECHNIQUES IN A SIMULTANEOUS–EQUATIONS CONTEXT

This section of the paper discusses the rationale for utilising robust estimation techniques as diagnostic tools in applied econometric research – particularly that research involving modelling macroeconomies in a time–series context. Many of the charges levelled against structural econometric modelling as a useful enterprise have been directed to the time–series context, rather than the cross–sectional context. The vulnerability of models estimated via standard least squares techniques over periods of well–accepted structural volatility to these critiques is obvious. As an alternative strategy to abandoning structural econometric modelling, considerable interest has been shown in developing more powerful and widely available diagnostics to aid the modeller and her critics to evaluate the degree to which the assumptions underlying classical modelling technques might be violated in each instance. Such well–known recent works as Belsley, Kuh and Welsch (1980) have raised researchers' consciousness in this area to a considerable degree.

Robust estimation techniques, as they might be applied to structural econometric modelling, are also a rather recent addition to the researcher's toolkit. Our emphasis in this study is to consider those techniques as analogous to regression diagnostics of the Belsley, et.al. type – not as alternative estimation techniques to be used in place of classical least squares techniques, but rather as adjuncts, used to diagnose problems with an estimated equation that might not be so readily visible otherwise.

These techniques were originated as a statistical approach to overcome deficiencies in the estimation of the location and scale parameter in univariate distributions. The regression model is only one of the possible generalizations of this framework, and it has not been fully investigated. For these and related reasons, including the unavailability of software to perform robust regressions, these methods have not been implemented in applied research. Monte Carlo studies that examine the performance of different robust estimators within the single equation model may be found in the literature, but it is very uncommon [illegible] ust applications [illegible] eal data. Robust regression has rarely been used as an estimation technique in single–equation econometrics, and for the simultaneous equation model the picture is even worse. The theoretical work in this area is very limited, in the sense that only a few robust estimators have been extended to the simultaneous equations model, and there appear to be no applications, either with real or with concocted data. The only simultaneous model for which there is a robust implementation uses the Least Absolute Deviation (LAD) estimator (Fair, 1984), but this method relies on a criterion of robustness different from that analyzed in this study.

This section is divided into two parts. The first provides some justification for this strategy, and argues against the quite inappropriate technique of "staring at the residuals." The second provides some analytical justification for the BIV estimation technique, and presents some analytical extensions to that technique due to Furno (1988a,1988b).

1.1 The Influence Function approach to regression diagnostics

Robust estimation techniques in econometrics are designed to lessen the dependence of the properties of the estimators on the basic assumptions of the classical regression model. The major goal is to obtain estimates that maintain good properties under small changes of the parametric model, that are more stable than least squares estimates, where stability implies that robust estimates are not critically dependent on the extreme values of the data set (outliers), which tend to perturb the least squares estimates.

These techniques consider the sample as a realization of a neighborhood of the assumed (or central) distribution, rather than regard the data as coming from the central distribution alone and as satisfying all the assumptions of the parametric model. The application of a bound on the residuals of the robust regression constrains those values that are "too far away" from the assumed distribution, so that the neighborhood considered collapses to the central distribution. To consider one realisation of this concept, assume that the distribution of the error term is $F_n = (1-\varepsilon) F_0 + \varepsilon H$, and while the assumption of zero mean and given variance still holds for F_0, nothing is known about the H distribution. The consequence is that, for the contaminated distribution F_n, the assumption $E(u)=0$ is unrealistic, and the bias is nonzero. Such a ... not vanish with the increase in the sample size, since the contamination is assumed to affect the sample in the constant proportion ε. If, rather than considering the single assumed distribution F_o, the analysis is extended to an entire neighborhood of F_o, an estimator is said to be consistent if $T_n \to T(F_o)$ in probability, and it is asymptotically normal for $\sqrt{n}\,[T_n - T(F_o)] \sim N(0, A(F_o, T_o))$.

To evaluate the concept of an estimator's robustness, we must define the influence function (IF). Roughly speaking the IF is the analogue of a derivative: the derivative of the statistic with respect to the degree of contamination in the data set, at the central distribution (Hampel 1974). Consider a sequence of estimators $\beta_n=\beta(F_n)$, where the estimator is a function of the distribution F_n. The latter is defined as the linear combination of the assumed function F_o and δ_x, the distribution giving a point mass of one to the observation x, $F_n \equiv (1-\varepsilon)\,F + \varepsilon\,\delta_x$. A degenerate distribution that assigns probability one to the single observation x (the outlier) is used to obtain a neighborhood of the central distribution, F_o. In the regression model, this allows to consider the impact of the single observation on the individual parameter estimate, and this is why such a model is called the infinitesimal approach. The influence function (IF) at the observation x is defined as:

$$(1) \qquad IF(x;\beta,F_o) \equiv \{\partial\,\beta(F_n)\,/\,\partial\varepsilon\}_{\varepsilon=0} = \lim_{\varepsilon\to 0}\,[\beta(F_n) - \beta(F_o)]\,/\,\varepsilon$$

where ε is the degree of contamination. It is possible to linearize the functional $\beta(F_n)$ in terms of $\beta(F_o)$, if β is sufficiently regular; this allows one to consider F_o as the main distribution of the model:

$$(2) \qquad \beta(F_n) = \beta(F_o) + \int IF(x;\beta,F_o)\,d(F_n - F_o) + \text{remainder}$$

This is just the analogue of a Taylor expansion, and says that the estimator at the underlying distribution F_n can be approximated by the value of the statistic at the assumed distribution F_o, plus the expectation of the first derivative of β with respect to the degree of contamination, plus a remainder. By definition, for ε=0, one has:

$$(3) \qquad \int IF(x;\beta,F_o)\,d(F_o) = 0$$

IF then provides an alternative method to consider the effects of the contamination on the OLS estimates. For $\varepsilon \rightarrow 0$, the term $n^{-1/2} \sum_i IF_i$ is asymptotically normal with mean zero and variance $V = E(IF)^2$. For $\varepsilon \neq 0$, however, $\int IF(x; \beta, F_0)\, dF_n \neq 0$, $\beta(F_n)$ is inconsistent. The bias does not vanish as the sample size goes to infinity, since the contamination is a given percentage ε of the sample size.

The derivation of robust estimators involves the definition of a bound on the influence function (IF) and thus on the residuals of the regression equation. This restriction on the residuals can be implemented through the inclusion of a weighting system; robust regression can be viewed as a particular kind of weighted least squares. The weighting system allows the researcher to single out observations that have a strong impact in the computation of the estimated parameters. The goal is to provide a good description of the bulk of data, forcing bad data to have large residuals. This is decidedly not the case with least squares, which attempts to bring those data points into the fold, and warps the hyperplane of best fit accordingly.

However, bounding the residuals is not costless, as it introduces nonlinearities in the estimation process, and causes a reduction in efficiency if the bound is erroneously imposed on well behaved error distributions. That is, if the assumed model is true, least squares estimates are more efficient than robust estimates, because of the Gauss–Markov theorem. However, the knowledge of the true underlying distribution is generally inaccurate, and the properties of the classical procedures (especially consistency) may well not be retained under discrepancies from the assumed distribution. The link between the two techniques (least squares and robust estimation) can be summarized in a tradeoff between bias and efficiency: extreme values in the matrix of the independent variables are likely to be gross errors and will cause bias; thus it is a good idea to downweight them. On the other hand, if they are not mismeasured values, the efficiency of the estimators is reduced by the downweighting, since a greater variance in the explanatory variables reduces the variance of the estimators.

The most common method used to detect outliers is the inspection of the residuals of the least squares regression. A large residual implies that the fitted value, $\hat{y}$, is far from the observed value y, and that the observation needs to be examined. This approach could be very misleading, as the concept of influential values indicates. Influential values are data points that have a particularly large role in computing the pa... ...imates. Their inclusion or exclusion in the estimation process substantially changes the estimated parameters, thus making least squares unstable. The line fitted via OLS is always very close to such influential observations, and the corresponding residual is small, not large.

Robust methods establish an objective criterion of intervention, which is particularly useful in large data sets, where the quality of the observations can not be individually verified. The weighting system they introduce is based on the maximum permissible residual, or, more generally, on the maximum allowed influence of each observation on the parameter estimates.

The comparison between least squares and robust results is helpful in detecting misspecification. If the standard assumptions of normality and finite variance are fulfilled, the point estimates of parameters obtained by least squares and those computed via robust methods should be quite close to one another. Of course, least squares estimates are preferable since, by the Gauss–Markov theorem, they are also the most efficient. If, however, the two estimators give very different results, this implies that the error process only approximately fulfills the theoretical assumptions. In this case, least squares estimates are inconsistent, and the robust regression is to be preferred.

1.2 Analytical foundations of robust simultaneous–equations estimation

We now present one of the different concepts of robustness, consider how it modifies the standard regression model, and derive the properties of the resulting estimator. The pioneering work in this field is due to Huber (1972) and Hampel (1971). A more recent estimator, proposed by Krasker and Welsch (1982), might be preferred; the latter two estimators are both based on the same concept of robustness, but the Krasker and Welsch estimator has recently been generalized to the simultaneous equation model (Krasker and Welsch,1985).

THE KRASKER–WELSCH BOUNDED–INFLUENCE ESTIMATOR

The bounded–influence estimator (BIF), presented by Krasker and Welsch (1982), relies on the concept of qualitative robustness. The main goal is to ensure that small departures from the assumed distribution cause only small changes in the estimates. The building block of this estimator is given by the influence function (IF), which measures the impact of each observation on the estimators, and by a statistic that is a transformation of IF, the gross error sensitivity, defined below. BIF minimizes the asymptotic variance of the regression while imposing a bound on the gross error sensitivity. The bound is chosen so that the asymptotic efficiency is only slightly impaired, and the estimator assumes an intermediate position in the trade off between unbiased and efficient estimators.

A comparison of the IF functions for different estimators illustrates the intuition behind the BIF estimator. For the linear equation $y=x\beta+u$, the IF of the OLS estimates is not bounded:

(4) $$IF_{OLS} = (y-x\beta)x^T / \int x^T x \, dF$$

This function can be increased by both large residuals and outliers in the x's. To better analyze the performance of the different estimators, it is helpful to split the IF in two terms: the influence of residuals, $IR= y-x\beta$, and the influence of position in the factor space, i.e. relative to the explanatory variables, $IP= x^T/\int x^T x \, dF$. OLS does not limit either of these terms, and is not robust in the Hampel sense. In contrast, the IF of the BIF estimator is given by

(5) $$IF_{BIF}=w(u,x)\,(y-x\beta)\,x^T / \int \partial/\partial\beta\,[w(u,x)\,(y-x\beta)\,x^T]\,dF$$

Here $w(u,x)$ defines the weighting system, which is a function not only of the residuals, but also of the x's. This allows one to control both IP and IR so as to bound the entire IF, where the goal is to neutralize the contamination and to bring the empirical distribution in line with the central one.

The objective function of the BIF estimator for the linear model is of the type:

(6) $$Q(\beta,\sigma) = n^{-1}\,\Sigma_i[w_i2(u_i, x_i)\,\rho(u_i/\sigma)+c]\,\sigma$$

where w_i is the weighting system that limits the impact of outliers in the estimation process. The parameters to be estimated are the coefficients β and the scale parameter σ, plus the parameter determining the weights. The first order conditions for β and σ are obtained [illegible] taking the partial derivatives. Estimation of the weighting parameter involves, first of all, the transformation of (5) to the form:

$$(7) \qquad IF_{BIF} = w(u, x)\,(y-x\beta)\,x^{T} / E\,[w(u, x)\,(y-x\beta)2/\sigma 2\; x^{T}x]$$

with asymptotic variance of the estimates of

$$(8) \qquad V_{\beta} = E(IF)^{2} = \sigma^{2}\,B^{-1}\,A\,B^{-1}$$

$$(9) \qquad B = E\,[w(u,x)\,(y-x\beta)^{2}/\sigma^{2}\;x^{T}x]$$

$$(10) \qquad A = E\,[w^{2}(u,x)\,(y-x\beta)^{2}/\sigma^{2}\;x^{T}x]$$

A can be interpreted as the standardizing factor for the x's, so that the estimator is invariant under linear transformations of x_{0}. The weighting system for the BIF estimator is derived by the imposition of a bound. In a regression with more than one explanatory variable, one has to impose the bound individually on the IF relative to each parameter of the equation. Alternatively, it can be considered as the impact of one observation on the entire set of parameters. This statistic is called the sensitivity function:

$$(11) \qquad r = \sup\{\, IF^{T}\,V_{\beta}^{-1}\,IF \,\}^{1/2} = \sup\, w(u,x)\,|(y-x\beta)/\sigma|\,(x\,A^{-1}\,x^{T})^{1/2}$$

The r statistic, a function of IF, may be shown to equal the absolute value of the standardized residuals, times a quadratic function called the robust distance:

$$(12) \qquad d(x) = (x\,A^{-1}\,x^{T})^{1/2}$$

Equation (11) is not the simple norm of the IF, as suggested by Hampel (1974), since invariance with respect to the coordinate system is required in the multivariate case. The value r must be invariant to both the units of measurement and the particular choice of explanatory variables, so that it is standardized by the asymptotic covariance matrix of the estimated parameters.

If the sensitivity function is finite, the IF is bounded, and vice versa. If $r \rightarrow \infty$, one can impose the bound a. The weights resulting from the introduction of the bound are

$$(13) \qquad w(u,d(x)) = \min\{\, 1,\; a/|t| \,\}$$

(14) $t = (u_i/\sigma)\, d(x) = (u_i/\sigma)\, (x^T A^{-1} x)^{1/2}$

The weights assume unit value for $r \leq a$. For large values of r, the weights attached to the corresponding observations are inversely related to the normalized residuals, and to the robust distance. As a rule of thumb, if the number of parameters of the model is k, a is suggested to be $(1.5\ \sqrt{k})$. In any case, a is chosen to exceed $\sqrt{k}$. The observations are weighted only if $r > a$ in order to minimize the loss of information, and of efficiency, involved in the downweighting process. Of course the loss in efficiency also depends also upon the bound a, since for $a \rightarrow \infty$ one obtains OLS, which are the most efficient estimates in absence of contamination. The smaller is a, the greater is the difference between OLS and BIF results, and the lower is the efficiency of the BIF estimates.

To complete the derivation of the BIF normal equations, consider equation (10). By replacing the weights with the explicit formula given by equation (13), one obtains

(15)
$$\begin{aligned} A &= E \min \{1\ ,\ a^2/[u_i^2/\sigma^2\ d^2(x)]\}\ u_i^2/\sigma^2\ x^T x \\ &= E \min \{u_i^2/\sigma^2\ ,\ a^2/d^2(x)\}\ x^T x \\ &= n^{-1} \Sigma_i\ w^2(u_i, d(x))\ x_i^2 \end{aligned}$$

To summarize, the weighting function derives from the introduction of a bound on the sensitivity r, which in turn depends upon the matrix A. The weights then are defined as a function of A and the final version of equation (15) gives the normal equation for w_i and A. The set of first order conditions is then:

(16) $n^{-1} \Sigma_i\ w_i(u, d(x))\ (y_i - x_i\beta)\ x_i^T = 0$

(17) $n^{-1} \Sigma_i\ w_i2(u, d(x))(y_i - x_i\beta)^2 = c\,\sigma$

(18) $n^{-1} \Sigma_i\ w_i2(u, d(x))\ x_i^2 = A$

where the constant c is defined as a function of the weights: $c = [^{(n-k)}/_n]\ \Sigma_i\ w_i(d(x)) = [^{(n-k)}/_n]\ E \min(1,\ a^2/d^2(x))$, that is, the average weight times the scale factor $[^{(n-k)}/_n]$. The model has a unique solution (β, σ, A), obtained iteratively, that is asymptotically normal and that converges "almost surely" to the true parameters (Maronna and Yohai, 1979).

The iterative solution is obtained through the following equations. The scale parameter is given by

(19) $\sigma 2_{m+1} = \Sigma_i \, (z_i)^2 \, / \, [((^{n-k})/_{n}) \, \Sigma_i w_i^2(d(x))]$

(20) $z_i = d(x_i)^{-1} \rho' \, [u_i/\sigma_m \, d(x_i)]\sigma_m$

(21) $d(x_i) = (n \, h_{ii})^{1/2} \, / \, w(u,d(x))$

The coefficients are updated through the equation $b_{m+1} = b_m + \tau$, where τ is given by

(22) $\tau = (x^T x)^{-1} \, x^T Z$

In equation (20), the Z_I are the winsorized residuals, that is residuals bounded by the function ρ', and in addition corrected for the presence of outliers in the x's by the term d(x), the robust distance. The latter in equation (21) is expressed in terms of h_{ii}, which is a key element in the detection of outliers. In the simple regression model, $y=x\beta+u$, the OLS estimates are given by $b=(x^Tx)^{-1}x^Ty$, and the predicted values of the dependent variable are $\hat{y}=xb=Hy$, where $H\equiv x(x^Tx)^{-1}x^T$ is the so called "hat matrix" which transforms y into its fitted values. H is a projection matrix, symmetric and idempotent, with diagonal elements satisfying the relations $0 \leq h_{ii} \leq 1$ and $tr(H) = \Sigma_i h_{ii} = k$, where k is the number of parameters in the regression (Huber, 1981). The diagonal elements of this matrix are particularly relevant as they measure the impact of y_i on $\hat{y}_i$, indeed $y_i - \hat{y}_i = (1-h_{ii}) \, y_i - \Sigma_{j \neq i} \, h_{ij} \, y_j$.

The latter relation shows how influential data can have small residuals. When $h_{ii} \approx 1$, a gross error (outlier) in y_i does not necessarily result in a large residual. Furthermore, if h_{ji} is large, the effect of a gross error in the ith observation will be reflected in the jth residual. This is why the OLS residuals are not reliable as a detection rule. The ideal situation is given by a balanced hat matrix, that is an H matrix with diagonal elements $h_{ii} \approx k/n$, and bounded–influence estimators, through the introduction of a weighting system, modify the observations to obtain such a condition. The diagonal elements of the hat matrix are used to define the weights attached to the observations. Consider again the robust distance as defined in equation (12). By squaring it, and by substituting out A, as defined in equation (18), one obtains

(23) $d(x_i) = (n \, h_{ii})^{1/2} \, / w_i(u,d(x))$

This is to say that the robust distance is a function of h_{ii}, which implies that the definition of the weighting system in the bounded–influence estimator depends on the elements that detect influential values. The weights themselves are an accurate diagnostic tool, preferable to the residuals of the OLS regression. The observations with low weights are the influential data in the sample.

THE BOUNDED–INFLUENCE INSTRUMENTAL VARIABLES ESTIMATOR

The Bounded–Influence estimator is one of the few robust estimators that have been extended to a simultaneous–equation context. The Bounded–Influence Instrumental Variables (BIV) estimator (Krasker and Welsch, 1985) is a quite straightforward extension of the single equation approach, with the inclusion of a set of instruments Q to solve the simultaneity problem. The assumptions about the instruments are the usual: $\operatorname{plim} n^{-1} Q^T X = M$, and $\operatorname{plim} n^{-1} Q^T U = 0$. The first order conditions are the following:

(24) $\quad 0 = \Sigma_i\, w_i(u,d(q))\, (y_i - x_i\beta)\, q^T$

(25) $\quad \sigma^2\, c = n^{-1}\, \Sigma_i\, w_i^2(u,d(q))\, (y_i - x_i\beta)^2$

(26) $\quad A = n^{-1}\, \Sigma_i\, w_i^2(u,d(q))\, q_i^T q_i$

Their derivation follows the derivation of the BIF estimator. The weights are $w_i = \min[1, a/\lambda_i]$, where the term λ_i is $\lambda_i = |u_i/\sigma|\, (q_i\, A^{-1} q_i^T)^{1/2}$, and the constant c in equation (25) is defined as:

(27) $\quad c = [{}^{(n-k)}/_{n^2}]\, w^2(d(q)) = [{}^{(n-k)}/_{n}]\, E \min[1\, ,\, a^2/d^2(q)]$

that is, the average weight times the factor ${}^{(n-k)}/_{n}$. The influence function, IF, for the case of a simple IV regression has an interesting insight:

(28) $\quad IF_{IV} = (y - x\beta)\, q^T / E\, q^T x$

It is worth noting that the above function can be unbounded only with respect to the residuals, since the instruments can be chosen so that the Q matrix is balanced. The x's enter the equation only through the determination of the residuals, and thus would be the only terms that need to be bounded. In contrast, the IF for the BIV estimator is given by:

(29) $\quad IF_{BIV} = w(u,q)\, (y - x\beta) q^T / E\, w(u,q)\, (y - x\beta)^2/\sigma^2\, q^T x$

The updating equation for the parameter vector is $\beta_{m+1} = \beta_m + \tau$, with

$$(30) \qquad \beta_m = (q^T wx)^{-1} q^T wy$$
$$(31) \qquad \tau = (q^T wx)^{-1} q^T wz$$

where z are the winsorized residuals. The updating formula of the variance is given in equation (25); the IV estimates are used as starting values.The asymptotic covariance matrix is computed as:

$$(32) \qquad A = n^{-1} \Sigma\, w^2\, q^T q$$
$$(33) \qquad B = n^{-1} \Sigma\, w_a\, q^T x$$

where $w_{ai}=1$ if $w_i = 1$ and $w_{ai} = 0$ otherwise. This new weighting system eliminates the downweighted observations: the data coming from the contaminating distribution are thus excluded from the computation of the variance of the estimated parameters.

The procedure outlined by Krasker and Welsch has only been available to applied researchers having access to MIT's TROLL system, in which a number of robust methods are provided. Since TROLL is not generally transportable outside the IBM (and plug–compatible) mainframe environment, the BIV procedure we use in this study was implemented by the authors in the well–known RATS™ econometric language. This makes BIV much more accessible, as RATS is available on many non–IBM computing systems (such as the DEC VAX, where our procedure was developed), as well as many microcomputers of the IBM–PC and Apple Macintosh families. A copy of the RATS procedure is available on request from the authors, in IBM PC or Macintosh floppy disk format; it also will soon be made available via dial–up access on the RATS Bulletin Board system with the proprietors' permission.

PROOF OF CONSISTENCY OF THE BIV ESTIMATOR WITH UNKNOWN σ^2

The usefulness of the Krasker and Welsch BIV estimator has been limited, since those authors were only able to establish (1985) that the BIV estimates are consistent and asymptotically normal under the restrictive assumption that the scale parameter of the equation, σ^2, is known. This assumption limits the validity of the estimator and the possibility of generalizing it to a three–stage procedure. In a forthcoming article, Furno (1988b) proves consistency of the estimated parameters (A_n, β_n, σ_n), with respect to underlying parameters (A_0, β_0, σ_0), where β_0's are the coefficients of the regression, σ_0^2 the scale parameter of the equation, A_0 the expectation of the weighted moment matrix of the instrumental variables, and n the sample size. This lends credibility to our use of the BIV technique as a regression diagnostic, since its analytical properties are now more clearly established, and should encourage its use in other contexts (such as an extension to a full–information setting).

2. APPLICATION OF BIV TO A CASE OF STRUCTURAL CHANGE

A common practice in applied econometrics is to introduce a dummy variable to deal with outliers, or a structural break, or any kind of sudden changes that appear to have occurred in the sample. This strategy has an obvious weakness: the dummy variable may neglect other phenomena related to a structural break. For instance, in a case of policy change, beyond the altered impact of a variable, there might well be a change in the stability of the regression, i.e. a change in the variance of the equation's error process, or of the variance of one or more regressors. We present one of these cases: a single equation that includes a dummy variable to model a policy change, which is part of a simultaneous model due to Fair (1984). The equation has been estimated both with and without the dummy variable to analyze whether such a variable models the structural change effectively.

A short–term interest rate equation is built on the assumption that the Federal Reserve has a target for the three–month Treasury bill rate, where the variables on the right hand side of the equation (real GNP and percentage change in the narrow money supply) define the reaction function. The structural change presumed to have occurred in 1979Q3 is the result of a Fed policy shift from interest rate targeting to money supply targets. This is modeled by the introduction of a multiplicative dummy for the money supply term (DD793) that assumes a value of zero before 1979Q3, and one afterwards, permitting a shift in the interest elasticity in the demand for money. The 2SLS estimates of the RS equation with the dummy variable, in equation 1A, and without it, in 1B, are given in Table 1. Table 2 reports the corresponding BIV results.

The RS equation is a particularly apt illustration of the relevance of diagnostics in a regression. If one considers Tables 1 and 2, there would be no doubt about the opportunity of introducing a dummy variable. The coefficient of the dummy is significantly [illegible] from zero both in 2SLS and BIV; its introduction reduces the estimated standard errors of regression, both in equations 1A and 2A; the sum of squared residuals (SSR) is lower in 1A and 2A than in 1B and 2B; the corrected R^2 is larger in the regression with the dummy; the residuals for the last period of the sample are larger in the regressions without the dummy, as can be seen in Figures 1 and 2. But there are two conflicting elements in the above picture: the Durbin–Watson statistic improves in the regression without the dummy; the residuals of the regression are small before 1979, and large afterwards. The latter is an imprecise signal of the persistence of a structural break in the equation, since, as stated earlier, the least squares residuals are not a very reliable diagnostic tool.

If we examine the weighting system introduced by the BIV regressions, a very clear pattern can be identified: very few observations are downweighted until 1970; some observations are downweighted from 1970 to 1979Q2, but the weights are quite small; almost all the observations are downweighted, and by a large extent, from 1979Q3 on. The series illustrated in Figure 3 are the weights of the BIV regressions, run with and without the dummy variable, for the 1975–1982 period. Both series have the same general pattern, which means that the introduction of the dummy can not be said to have solved the problem. But what is surprising is that the introduction of the dummy has worsened the picture: equation 2A has a larger number of observations downweighted, with a sum of weights of 108.61, versus 111.7 in equation 2B, and 115 in equation 1A. The weights are generally smaller in the case with the dummy, indicating that the equation with the structural shift actually requires more modification to generate the robust estimates.

The message carried by the BIV weights is quite clear. Although both economic theory and standard econometric tests ("Chow tests") would suggest that a structural change in the interest elasticity of the demand for money should be implemented in this equation, the BIV results point out the flaw in that strategy: the estimated equation with the dummy still requires sizable downweighting to derive the robust estimates. The econometric problem which BIV is pointing out is, indeed, the maintained hypothesis that the error variance has not changed in the post–1979 period. A set of tests of that hypothesis have been conducted. By dividing the residuals into two subsamples, until 1979Q3 and afterward, one can perform an F–test to verify the null hypothesis that the error variance has not increased after the policy change. The F has been computed in the four cases, with and without dummy, using BIV and 2SLS, and is presented in

...bles 1 and 2 as "F". As one would expect, the computed F's are larger in the regressions without the dummy. However, the values in the equations with the structural shift still greatly exceed the critical value: the hypothesis of an unchanged error variance can not be supported. Furthermore, the F's of the robust regressions are larger than the corresponding F's of 2SLS, thus the rejection of the null hypothesis is much stronger in the BIV case. In this sense, robust BIV estimation may not only help the applied researcher to diagnose the existence of a problem in an estimated equation, but may also provide a stronger statistical test. Since BIV estimation is not as heavily oriented toward reducing large residuals as is least squares, the large residuals that would help confirm an unstable error variance are more visible in the BIV than in the 2SLS case. Thus this example—using an estimated equation from Fair (1984)—demonstrates the degree to which BIV estimation may have considerable diagnostic potential in a single–equation context.

3. Application of BIV Estimation to Klein's Model I

This section presents the BIV estimation of a small macroeconometric model, Klein's well–known "Model I" (1950), and contrasts the resulting estimates with their least squares counterparts. The BIV and two–stage least squares estimates are compared in terms of both their effects on the individual equations' behavior, and their effects on the performance of the model as a whole.

Klein's Model I of the U.S. economy includes a consumption function, with current and lagged profits and the total wage bill as explanatory variables; an investment equation, where the right hand side variables are given by profits – current and lagged – and the stock of productive capital; and a demand for labor equation, driven by the wage bill paid in the private sector, which is a function of total production in the private sector – current and lagged one period – and a time trend, which is intended to capture the increasing strength of unions in the interwar period. Three identities close the model, which is estimated from annual data for the interwar period.

As a first illustration of the diagnostic use of BIV for this model, the Private Wage (demand for labor) equation has been estimated with the bound varying between 2.0 – a very tight constraint – to 4.0, a relatively loose constraint, where the BIV results are essentially those of the least squares method. The suggested bound for this equation, using the rule of thumb $1.5 \sqrt{k}$, is 3.0. A summary of the key results from several applications of the BIV estimator is presented in Table 3.

The BIV estimates of this equation are generally less precise than those of 2SLS, as judged by the standard error of estimate. Even a tight bound produces relatively little downweighting from the unbounded sum–of–weights of 21 (equal to the number of observations). A standard diagnostic technique proposed by Denby and Mallows (1977) involves plotting the computed residuals and/or the estimated coefficients from several values of the bound. As the bound is tightened, observations with increasing residuals are those which heavily influence the least squares results; coefficients which vary are those which are unstable, or heavily influenced by certain data values. Figures 4 and 5 present the residuals and estimated coefficients resulting from BIV estimation with bounds between 2.0 and 4.0. They suggest that this equation is relatively stable, in that few observations' residuals vary noticeably between tightly–bounded and effectively unbounded estimates. The coefficients, as well, are seen to be quite stable; only the first coefficient (the constant term, which is not significantly different from zero) seems to move as the bound is tightened.

The variation in the estimated standard error of estimate between the essentially unbounded BIV result (with a bound of 4.0) and the 2SLS result might be questioned. It must be noted that even if there is no change in the estimated coefficients, and there is no downweighting, the standard error of the regression and, thus, the standard errors of the estimated coefficients are larger in the BIV case. This is a clear example of the tradeoff existing in robust techniques between bias and efficiency: in the case of a unit weights vector, when there is no gain in terms of bias in the estimated coefficients, the variance of the regression error (and of the estimated coefficients) of necessity increases. It follows that in these cases the t–tests have a greater rejection power: the coefficients that are barely significant in 2SLS will turn out not significantly different from zero in BIV. The estimated standard error of regression in BIV does not coincide with the 2SLS estimate because of the downweighting that BIV performs on the instrumental variables. Even if the final weights are all unit elements, there are observations in the set of instrumental variables that are downweighted. These downweights enter in the computation of the variance of the regression through the constant *c*, in the equation for the BIV variance:

(34) $\sigma^2 c = n^{-1} \Sigma_i w_i^2 u_i^2$

(35) $c = [^{(n-k)}/_n]\, E \min[1, a^2/d^2(q)]$

In equation (35), as long as the term $a^2/d^2(q)$ is smaller than one, the constant c is smaller than $^{(n-k)}/_n$, and the estimated variance in BIV is larger than its 2SLS counterpart.

SIMULATION RESULTS

The performance of BIV estimators might also be examined through the analysis of model simulation statistics. In this context, all behavioral equations of a model are estimated via BIV, and the resulting point estimates used in a nonstochastic simulation, which may then be compared with the traditional 2SLS–based counterpart. To permit the calculation of both in–sample and out–of–sample forecast statistics for the model's equations, the estimation period for Klein's model was reduced to 1921–1938, with the three years 1939–1941 reserved for *ex ante* simulations. Table 4 presents the standard error of estimate, out–of–sample root mean squared error (RMSE), and out–of–sample mean absolute error (MAE) for the three estimated equations. In all three estimated equations, the 2SLS estimates possess better in–sample qualities than the BIV counterparts; this ranking also holds for the investment function in terms of *ex ante* forecasting performance. In the investment equation, where the BIV regression with a bound of 3.0 does not impose any downweighting, the simulation results from BIV and 2SLS coincide. However, one should note that the BIV estimates of the consumption function produce marginally better out–of–sample performance, and those of the private wage function are even stronger – despite the weaker in–sample statistics of those estimates. Thus, it should be noted that if *ex ante* forecasts are one of the objectives of modelling, BIV estimates may have something to offer; their estimated coefficients are less sensitive to the peculiarities of the estimation period, and thus may be likely to provide at least a useful check on the 2SLS results. In this sense, we would propose that a BIV–estimated model might be used to benchmark the forecasts of a 2SLS–based model, and notable divergences utilised to analyse defects in the central modelling strategy.

This "sanity check" is performed for Klein's Model I in the context of a static simulation over the 1939–1941 period, using, respectively, 2SLS and BIV coefficients estimated over the 1921–1938 range. BIV bounds of both 2.0 and 3.0 were used to illustrate the sensitivity of the model's simulation performance to this "tuning parameter". The results for the six endogenous variables of Klein's model are presented in Table 5. We see that the BIV results for either bound are generally weaker than their 2SLS counterparts, with the exception of the private wage (PW) equation. The satisfactory performance of 2SLS in this regard gives credence to the belief that this model is reasonably well estimated with standard 2SLS techniques; the econometric problems which might allow robust techniques such as BIV to outperform least squares do not seem to be present. This is far from an uninteresting conclusion to this examination of Klein's model, as it is analogous to a "clean bill of health". We would propose that BIV estimation and simulation might well be used in this manner to ensure that, indeed, there are no meaningful discrepancies between BIV and classical techniques' results.

4. APPLICATION OF BIV ESTIMATION TO FAIR'S LINUS MODEL

This section analyzes the performance of the Bounded–Influence Instrumental Variables robust estimation method in the context of another simultaneous equations system, more sizable (in terms of both numbers of equations and observations) than Klein's model. This greater size should ensure the presence of some outliers and gross errors, and would allow BIV's capabilities to be examined in a not–so–well behaved sample. The model that will be estimated is a smaller version of the well–known macroeconomic model of the United States economy presented by Fair (1984). Fair's large model consists of 128 equations, some involving non–linearities, and the non–linearity causes severe estimation problems for robust techniques, since in the case of the BIV estimator linearity is a required assumption.

A 12 linear equation system of the US economy, the so called LINUS model (Fair,1984), has been chosen. Fair, in comparing the performance of LINUS with the larger model finds that, in terms of predictive accuracy, the MAE and RMSE yield less satisfactory results in the small model. With the caveat that the use of a smaller version of a model may worsen its capability to explain the behavior of the economy, LINUS still represents a good framework to test different procedures. LINUS is essentially an IS–LM model, with consumption, investment, production and a monetary sector. The first three equations are the consumption functions for services (CS), nondurables (CN), and durables (CD); in each equation consumption depends on its respective lagged value, real gross national product (GNPR), and an interest rate variable. The short–

term interest rate (RS) enters the equations of consumption of services and nondurables, while the long–term rate (RM) is included in the equation of durable goods consumption. Inclusion of gross national product in these three equations is intended to reflect the conditions of the labor market; in periods of recession the wage rate is correlated with income, which proxies a "demand pressure" measure.

Investment is divided into housing and non–housing categories. The demand for housing (IHH) is a function of the long–term rate, gross national product, and housing investment lagged one period; non–housing investment (IKF) is a function of the actual and lagged level of production (Y), the depreciation of capital (KK), and the lagged investment. The aggregate production function (Y) depends on three assumptions: i) inventories are proportional to current sales; ii) production equates sales and a fraction of the gap between desired and actual inventories; iii) actual production partially adjusts to desired production. Thus Y is a function of its lagged value, total sales (X), and the stock of inventories in the previous period (V).

The mortgage rate equation (RM) is a term structure equation: following the expectations theory, the long–term rate is a function of current and future values of the short–term rate. In the estimation, the unobservable future values are proxied by the current and past values of the short–term interest rate. The short–term interest rate equation (as presented above in Section 2) is built on the assumption that the Federal Reserve has an interest rate target each period, and uses as policy instruments the variables on the right hand side of the equation. Following the shift in the Federal Reserve operating policy that occurred on 1979Q3, the relevance of money supply growth increases; this is modeled by the introduction of a multiplicative dummy for the money supply term that assumes a value of zero before 1979:3, and one afterwards. Four identities close the system.

The data are quarterly observations from the first quarter of 1954 to the third quarter of 1982; the variables of the model are listed in Table 6. A more detailed description of the sources and of the transformations involved in the determination of the data set can be found in Fair's Appendix A (1984). Since Fair published his book, many variables in the data set have been revised. This implies that, even using Fair's sample period, the coefficient estimates based on the revised series do not quite coincide his original values. The fact that, after only a few years, it is impossible to exactly replicate Fair's results would justify the choice of a robust procedure:one which offers greater

...ity with respect to small changes in the data. Of course robust methods do not offer any guarantee against extensive revisions, that is, large changes in the majority of data points.

The set of first–stage regressors used in each equation to compute the fitted values of the right hand side endogenous variables includes a constant term, CD_{-1}, CD_{-2}, CN_{-1}, CN_{-2}, CS_{-1}, CS_{-2} , $(DD793 \cdot M1)_{-1}$, $(DD793 \cdot M1)_{-2}$, $GNPR_{-1}$, $GNPR_{-2}$, IHH_{-1}, IHH_{-2} , IKF_{-1}, IKF_{-2}, KK_{-1}, KK_{-2}, $M1_{-1}$, $M1_{-2}$, Q_1, Q_2, RM_{-1}, RM_{-2}, RS_{-1}, RS_{-2}, RS_{-3}, V_{-1}, V_{-2}, Y_{-1}, and Y_{-2} (the same set of instruments used by Fair). The choice is not arbitrary, but is guided by the necessity of correcting a first order autocorrelation process in four of the eight stochastic equations of the model: the consumption equations for nondurables, for services, the production function, and the housing investment equation. Fair (1970) shows that the set of instruments required in a simultaneous system with lagged endogenous variables and first order autocorrelation must include all the predetermined variables of the system, as well as the endogenous and the predetermined variables lagged one period.

Robust methods rest on the assumption of independent and identically distributed errors (i.i.d.), and thus require the elimination of serial correlation. Since outliers or departures from normality affect the estimation of the autocorrelation coefficient, ρ, in the same manner that they affect the other parameter estimates, a robust estimate of ρ would be in order. This would be easily done by constructing robust estimates of the equations of the model, but the t–statistics of the estimated coefficients are not as reliable in the robust regression as in the least squares case, and this implies that the autocorrelation assumption is not directly testable. Indeed, the validity of the t–test is weakened by the assumptions upon which robustness relies: if the empirical distribution of the error term is given by the mixture of two different distributions, then the distributions of the statistics that are function of the residuals hold only with some approximation. The weighting system introduced by a robust regression tries to purge the effect of the contaminating distribution; but since the degree of contamination is unknown, the procedure is weakened and this reduces the power of the tests. This is why a robust estimate of the autocorrelation coefficients has not been performed; rather, the 2SLS estimates of ρ have been used to transform the data before implementing the robust regression.

The estimates of the eight stochastic equations of the LINUS model for the 1954Q1–1982Q3 sample period are given in Tables 7.1–7.8. For purposes of comparison, both the 2SLS estimates (as reestimated from currently available data) and the BIV estimates are shown. The robust estimates are computed by defining the bound equal to $1.5\sqrt{k}$ (where k is the number of parameters to be estimated). This ranges from 3 to 3.6, according to the number of coefficients in each equation.

Vis–a–vis the 2SLS results, there are two changes in sign: the constant term in the housing investment equation, and the RS_{-1} coefficient in the mortgage rate equation; both coefficient estimates have very low t–statistics in both cases. All the BIV coefficient estimates fall within the two standard error confidence interval of the 2SLS values. Furthermore, the RM_{-1} coefficient in the CD equation, the constant term in the housing investment equation, the Y_{-1} coefficient in the production equation, the constant, the GNPR and the $GNPR_{-1}$ in the Treasury bill equation, and the RS_{-1} and the RS_{-2} coefficients in the mortgage rate equation all have estimated coefficients that are not significantly different from zero in the BIV regressions.

More generally, it is possible to single out the following features of the BIV results:

i) The 2SLS results are confirmed by the BIV method in the CS, CN, IHH, and IKF equations.

ii) The parameter estimates with low t–statistics obtained using 2SLS become smaller in absolute value with larger standard errors in the BIV case.

iii) The BIV estimator produces an increase in the coefficients of the lagged endogenous variables, unless these coefficients are not significantly different from zero.

iv) The widest discrepancies between 2SLS and BIV results appear in the CD, Y, RS and RM equations.

v) The estimated standard errors of the BIV regressions are always smaller than the ones in the 2SLS.

ANALYSIS OF THE STANDARD ERRORS OF REGRESSION

As mentioned, in well–behaved error distributions the least squares estimator gives the best (i.e. minimum variance) estimates. A reduction of the robustly estimated standard error of estimate, which takes place in all the equations of the model, points out that there are some assumption of the Gauss Markov theorem that are not fulfilled, and that worsen the least squares performance. Such reduction would support the use of robust regression. The decrease of the robust versus the least squares standard error is the result of the downweighting of the error terms implied by the BIV etimators. Thus, the greater the reduction in the standard error of the regression, the worse the performance of the least squares estimator, the greater the downweighting or the bounding of the error terms performed by the robust regressions. The robust standard error can then be considered as a synthetic measure of the weighting system implemented on each equation.

Table 8 compares the behavior of the standard errors of the different techniques. The standard errors of the 2SLS and BIV estimators are presented in columns (1) and (2). Column (3) reports the percentage reduction in the BIV estimator standard errors with respect to their 2SLS counterparts. The RM, RS and CD equations present the largest reductions in their standard errors. The fact that these equations display such a large reduction, is again a signal of the less than satisfactory 2SLS estimates of the interest–rate and durable goods consumption equations.

To have a more detailed picture about the effects of the BIV estimator, one can analyze the weights it produces. Considering the entire LINUS model, one may note that the majority of downweighting is concentrated in the last period of the sample, during the 1970's and the beginning of the 1980's, while very few observations are downweighted in the period 1954Q1–1969Q4. The fact that the weights hardly ever come into play before the 1970's can be linked to the greater stability of the economy before the oil crisis and an increase in the variances of the model in the most recent years. This could be modeled by the introduction of a structural break or of non–linearity in the equations (as mentioned, with respect to the RS equation, in Section 2 above). It may very well be the case that the downweighting occurs at the end of the sample to force linearity upon what in reality is a curvature of the function, or a shift in the underlying economic structure.

In terms of weights on the instrumental variables, which are function of the robust distances, one can consider the sum of the weights (SW) as an index of the downweighting performed by the BIV estimator. The maximum attainable value of SW is given by the sample size and corresponds to the case of unit weights for all the observations, [illegible] this case SW_{max}=[illegible], which [illegible] [illegible] case. In the model the highest SW values, i.e. the weighting system closest to 2SLS, are found in the Y and IKF equations, respectively SW_Y=113.27 and SW_{IKF}=113.33; the lowest SW value, indicating extensive downweighting, is reached by the RS equation, SW_{RS}=105.07, followed by the RM equation, SW_{RM}=107.35, by the IHH and CD equations, SW_{IHH}=109.54, SW_{CD}=109.98, and finally by SW_{CS}=111.88, SW_{CN}=111.68. The large value of SW in the Y equation confirms the hypothesis of a presence of large residuals in the Y equation coupled with small robust distances.

The weights due to the robust distances (due to outliers in the explanatory variables) do not necessarely coincide with the final weights, since the latter are the resultant of both robust distances and residuals. In terms of final weights, the largest number of weights smaller than unity is found in the RM and IHH equation (25) followed by the RS equation (24) and the CD equation (23). The observations downweighted in the CN and IKF equation each total 21, while there are 20 in the Y equation. The CS equation is the one characterized by the least intervention, with only 16 weights less than unity. What is perhaps more interesting is the number of final weights substantially smaller than unity, for instance smaller than 0.7. The latter value is, of course arbitrary, but it gives a benchmark to evaluate the intensity of the downweighting. The RS equation is the one most heavily downweighted, where 41% of the weights different from unity are lower than the 70% benchmark. The RM equation follows with 40%, the CD equation with 21%, CS with 18%, and so forth, up to the CN equation that has no weight as small as the chosen benchmark of 70%.

ANALYSIS OF SINGLE–EQUATION SIMULATIONS OF THE LINUS MODEL

The analysis of simulation results is usually used to compare the predictive ability of different models — that is, how close they can reproduce the economy's behavior. In this context it can be used to test statistical methods instead of models, since the comparison of simulations can be considered as a criterion to evaluate the goodness–of–fit of the entire model under the different estimating procedures. Furthermore, outliers are among the major factors determining poor forecasts. Thus, it can be possible to verify if (and by what extent) prediction errors are the result of inappropriate statistical techniques – techniques that do not take into account the presence of outliers. The comparison between 2SLS and robust predictions can offer additional insights into analyzing the applicability of robust methods.

Since BIV is a single equation estimation method, it may be interesting to look at the forecasts of each single equation before analyzing the simulations of the entire model. Table 9 presents the results of one-step-ahead, in-sample forecasts, performed individually on each equation of the model using both the 2SLS and BIV estimates. The statistics that summarize the simulation results are the RMSE and MAE. The MAE statistic is less sensitive to large forecast errors, and thus may be considered the most appropriate statistic to assess the performance of the BIV estimator. The minimization of the bounded residuals performed by BIV improves the forecasts in the majority of the observations; but in case of influential outliers, the BIV objective function implies by construction that the fitted values of the robust regression are far from the observed values, generating large forecast errors. The latter may have an exceedingly large impact on the RMSE statistic.

The in-sample results of Table 9 show an improvement of the MAE values in all the BIV equations vis-a-vis 2SLS counterparts, except for the Y equation. The RM variable shows a considerable improvement in the forecasts, followed by IHH, CD and RS. In terms of RMSE, the results are not as favorable to the BIV estimates, but, as aforementioned, there are reasons to believe the MAE to be a more accurate statistical index. Regardless, the extent of the worsening of the BIV versus the 2SLS RMSE is rather small in most of the cases.

ANALYSIS OF SIMULATIONS OF THE ENTIRE LINUS MODEL

We turn now to the comparison of simulations of the entire LINUS model via 2SLS and BIV, both over the sample period and for sixteen additional quarters. Fair's model is estimated over the period 1954Q1 to 1982Q3, so that the data for the latest years, 1982Q4 to 1986Q3, are available to verify the out-of-sample forecasts of the model. These are static, deterministic simulations, which take into account the four additional variables defined in the identities of the model. Table 10 presents the RMSE and MAE values of these simulations for the sample period; Table 11 presents corresponding values for the *ex ante* simulations over 1982Q4–1986Q3.

The in-sample comparison between the 2SLS and the BIV shows that most of the variables have a smaller RMSE in the robust regression case, and all the variables but

Y, and thus GNPR and V, have MAE smal[illegible] [illegible]LS case. The GNPR variable is defined as GNPR=$Y+Q_2$, so that a bad performance in Y is bound to be reflected in GNPR. The same is true for V, defined as $V=V_{-1}+Y-X$ (columns (5) and (6) in Table 10).

For the out-of-sample simulations the BIV estimator generally gives smaller RMSE than 2SLS, for all but the IHH and Y, and consequently the GNPR and V, variables (columns (5) and (6) in Table 11). An interesting characteristic of these simulations must be pointed out. In the RS, CD and RM variables, the reduction in the RMSE of the robust versus the 2SLS model is very sizable; in the BIV case such reductions are, respectively, 17, 14 and 11 per cent. The equations describing these three variables were the ones showing the greatest divergence between the robust and the 2SLS techniques in terms of the coefficient estimates and the largest reductions in the standard errors of regression.

In the simulations of the LINUS model, we may distinguish three groups of variables:

i) The CS, CN, CD, IKF, RS, RM, X, and KK BIV forecasts improve in both the in-sample and the out-of-sample case.
ii) Y, GNPR and V worsen in both cases for the BIV forecasts.
iii) The IHH BIV forecasts improve in-sample and worsen out-of-sample.

Thus one can conclude that the results of the comparison of the simulations confirm an improvement in forecasting accuracy of the robust techniques versus their 2SLS counterparts in most of the variables. This said, we would again suggest that the particular value of the BIV technique might be as a diagnostic tool; its ability to flag observations having particular impact on the least squares results may well be of value to the modeller, since divergence of 2SLS and BIV simulation results might also point up a weakness in model specification.

5. Conclusions and Suggestions for Further Research

In this paper, we have described an estimation technique – bounded-influence instrumental variables estimation – which may be profitably used to diagnose time-series regression equations of the sort commonly used for econometric modelling and policy analysis. Although the technique is not a new one, we have presented both analytical results on its consistency and an empirical implementation via a widely-used econometric software package that should make it more appealing to the applied researcher.

The application of BIV to Fair's short-term interest rate equation has served to illustrate the degree to which commonly applied diagnoses of (and remedies for) structural change may not be appropriate. Summary statistics and plots based on BIV estimates may have considerable power in such cases to indicate the degree to which classical assumptions are satisfied – even in the absence of a large sample. This may be of particular interest to researchers working with the relatively few time-series data points generated in the aftermath of deregulation, tax reform, and similar institutional changes in the macroeconomic environment.

The analysis of Klein's Model I and Fair's LINUS model via BIV techniques has illustrated how these single-equation techniques might profitably be employed in the specification and validation of a simultaneous macroeconometric model. The BIV estimates of coefficients, and the associated *ex ante* predictions, may be used to explore the robustness of classical estimates in the face of structural instabilities — especially as they appear in the context of model simulation, in which prediction errors in one equation are propagated through the simultaneous system.

Additional research is underway to explore the degree to which a BIV analogue might be constructed for systems estimators. Since many simultaneous-equations models may be constructed with full information techniques, it would be very useful to have a bounded-influence three-stage estimator, or a "BFIML". This would be of particular importance, given the fragility of systems estimates to misspecification across equations. Additional work is also needed to develop BIV implementation procedures so that the estimators presented here may be calculated with no more complexity than, say, AR(1) estimates. This involves not only streamlining the user interface to a BIV implementation, but also improving its internal logic to quicken convergence and minimise temporary storage requirements. Thought should also be given to design of the most valuable graphical representations of BIV. As illustrated in our two examples, graphical presentation of, for example, BIV downweighting may be much more communicative that a tabular listing or summary statistic. As more econometrics packages are available on microcomputers with powerful graphical interfaces, this communication should be enhanced.

REFERENCES

Belsley, D., Kuh, E., and R. Welsch, 1980. *Regression Diagnostics*. Wiley, New York.

Denby, L. and C.L. Mallows, 1977. Two diagnostic displays for robust regression analysis. *Technometrics*, 19:1–13.

Fair. R., 1984. *Specification, Estimation and Analysis of Macroeconomics Models*, Harvard University Press.

Furno, M., 1988a. *Robust Methods for Macroeconomic Models.* Unpublished Ph.D. dissertation, Department of Economics, Boston College, Chestnut Hill, MA.

__________, 1988b. Bounded–Influence Instrumental Variables Estimator: an Extension. *Economics Letters*, forthcoming.

Hampel, F., 1971. A General Qualitative Definition of Robustness. *Annals of Mathematical Statistics*, vol. 42, No. 6.

__________, 1974. The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association*, 69.

Huber, P., 1972. Robust Statistics: a Review. *Annals of Mathematical Statistics,* vol. 43, No 4.

__________, 1981. *Robust Statistics.* John Wiley, New York.

Klein, L., 1950. *Economic Fluctuations in the U.S., 1921–1941.* Cowles Commission Monograph no. 11.

Krasker, W., and R. Welsch, 1982. Efficient Bounded–Influence Regression Estimation. *Journal of the American Statistical Association*, 77.

__________, 1985. Resistant Estimation for Simultaneous Equations Models Using Weighted Instrumental Variables. *Econometrica*, 53, No 6.

Maronna, A., H. Bustos, and J. Yohai, 1979. Bias and Efficiency–Robustness of General M–Estimators for Regression with Random Carriers, in *Smoothing Techniques for Curve Estimation*, T. Gasser and M. Rosenblatt eds, Lecture Notes in Mathematics No.757, Springer, Berlin.

Table 1: Treasury bill rate, 2SLS results

1A) $RS = -.152 + .892\ RS_{-1} + .01012\ GNPR - .00999\ GNPR_{-1} + .05044\ M1_{-1}$

(.379) (.05542) (.00296) (.00307) (.0217)

$+.142\ DD793{\cdot}M1_{-1}$

(.0335)

SE =.723, R^2 =.951, DW = 1.62, F = 6.4009, SW = 115

1B) $RS = -.0426 + .99939 RS_{-1} + .0102\ GNPR - .01042 GNPR_{-1} + .08668\ M1_{-1}$

(.40) (.05) (.0031) (.0033) (.021)

SE = .778, R^2 = .94, DW = 1.73, F = 10.005, SW = 115

Note: Standard errors in parentheses; the sample period is 1954Q1–1982Q3.

Table 2: Treasury bill rate, BIV results

2A) $RS = -.1426 + .913\ RS_{-1} + .00521 GNPR - .005083 GNPR_{-1} + .04919\ M1_{-1}$

(.299) (.0610) (.00319) (.00329) (.0176)

$+.1073\ DD793{\cdot}M1_{-1}$

(.0798)

SE = .4872 F = 9.8035 SW = 108.61

2B) $RS = -.00486 + .9927\ RS_{-1} + .00605\ GNPR - .006175\ GNPR_{-1} + .06709\ M1_{-1}$

(.30) (.05) (.0034) (.0035) (.017)

SE = .515 F = 13.5628 SW = 111.70

Note: Standard errors in parentheses; the sample period is 1954Q1–1982Q3.

Table 3: BIV Estimates of Klein's [illegible]ion, 1921–1941

Bound	Sum of Weights	Std.Err.Estimate
2.0	19.540	0.9519
2.5	20.741	0.8061
3.0	20.937	0.8360
3.5	20.974	0.8454
4.0	21.000	0.8526
∞	21.000	0.7672

Note: a bound of ∞ refers to the 2SLS estimates.

Table 4: Forecasting Accuracy of Klein's Behavioral Equations

	Std.Err.Estimate	RMSE	MAE
C–2SLS	0.8454	2.189	1.399
C–BIV	0.9558	2.181	1.395
I–2SLS	1.0301	1.199	1.198
I–BIV	1.1799	1.199	1.198
PW–2SLS	0.7605	0.930	0.754
PW–BIV	0.8502	0.907	0.751

Notes: C, I, PW refer to the consumption, investment, and private wage equations, respectively. Std.Err.Estimate is calculated from the estimation sample 1921–1938. RMSE and MAE are calculated from the out–of–sample observations 1939–1941. All BIV estimates are computed with a bound of 3.0.

Table 5: Out-of-sample Klein's Model I simulations, 1939–194[illegible]

	2SLS, a=∞		BIV, a=3		BIV, a=2	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
C	2.60	1.60	2.60	1.60	2.56	1.62
I	2.40	1.83	2.44	1.87	2.76	2.20
PW	2.43	2.14	2.40	2.11	2.24	1.96
Y	4.94	3.24	4.98	3.27	5.22	3.52
K	2.40	1.83	2.44	1.87	2.76	2.20
Π	2.85	2.09	2.90	2.09	3.22	2.13

Notes: Y, K, and Π refer to production, the capital stock, and corporate profits, respectively, which are defined in the three identities closing Klein's model.

Table 6: List of LINUS Model Variables

CD	consumer expenditures for durable goods
CN	consumer expenditures for nondurable goods
CS	consumer expenditures for services
DD793	0 until 1979Q2; 1 afterward (dummy for Fed policy shift)
dk	depreciation rate of capital, equal to .0263
GNPR	real gross national product
IHH	residential investment, households
IKF	nonresidential fixed investment
KK	stock of capital
M1	percentage change in the money supply
RM	mortgage rate, percentage points
RS	three month bill rate, percent
V	stock of inventories
X	total sales
Y	production

Table 7.1: <u>Consumption of services</u>

$$CS = -1.438 + .978\ CS_{-1} + .00332\ GNPR - .220\ RS$$
$$\quad (.504) \quad (.0105) \quad (.000923) \quad (.0366)$$
$$SE=.746,\ R2=.999,\ DW=2.15,\ \rho=-.258$$
$$(.092)$$

2SLS estimates

$$CS = -1.454 + .978\ CS_{-1} + .00331\ GNPR - .207\ RS$$
$$(.4326) \quad (.00914) \quad (.00079) \quad (.03905)$$
$$SE = .5977$$

BIV estimator results

Table 7.2: <u>Consumption of nondurables</u>

$$CN = 5.462 + .789\ CN_{-1} + .0120\ GNPR - .298\ RS_{-1}$$
$$(1.86) \quad (.0632) \quad (.0033) \quad (.0718)$$
$$SE=1.033,\ R2=.998,\ DW=1.94,\ \rho=.259$$
$$(.11)$$

2SLS estimates

$$CN = 5.1678 + .798\ CN_{-1} + .0116\ GNPR - .297\ RS_{-1}$$
$$(1.742) \quad (.0614) \quad (.00332) \quad (.0919)$$
$$SE = .9294$$

BIV estimator results

Table 7.3: <u>Consumption of durables</u>

$$CD = -4.305 + .789\ CD_{-1} + .00663\ GNPR - .264\ RM_{-1}$$
$$(1.202) \quad (.0524) \quad (.001507) \quad (.0933)$$
$$SE=1.507,\ R2=.991,\ DW=1.97$$

2SLS estimates

$$CD = -3.537 + .832\ CD_{-1} + .00527\ GNPR - .176\ RM_{-1}$$
$$(1.064) \quad (.0492) \quad (.00131) \quad (.1497)$$
$$SE = 1.1111$$

BIV estimator results

Table 7.4: Housing investment

$$IHH = -1.034 + .320\ IHH_{-1} + .0146\ GNPR - 1.699\ RM_{-1}$$
$$\quad (3.90) \quad (.0831) \quad (.0022) \quad (.237)$$
$$SE=1.085,\ R2=.97,\ DW=2.202,\ \rho=.892$$
$$(.051)$$

2SLS estimates

$$IHH = .460 + .314\ IHH_{-1} + .0141\ GNPR - 1.701\ RM_{-1}$$
$$(3.603) \quad (.1212) \quad (.00333) \quad (.435)$$
$$SE = .9149$$

BIV estimator results

Table 7.5: Production

$$Y = -8.323 + .127\ Y_{-1} + 1.036\ X - .1600\ V_{-1}$$
$$(2.67) \quad (.0412) \quad (.0438) \quad (.0307)$$
$$SE=2.45,\ R2=.999,\ DW=2.203,\ \rho=.523$$
$$(.090)$$

2SLS estimates

$$Y = -9.59 + .0539\ Y_{-1} + 1.1163\ X - .1644\ V_{-1}$$
$$(2.752) \quad (.0651) \quad (.0667) \quad (.0348)$$
$$SE = 2.366$$

BIV estimator results

Table 7.6: Investment

$$IKF = -2.66 + .935\ IKF_{-1} - .00675\ KK_{-1} + .123\ Y - .0955\ Y_{-1}$$
$$(.658) \quad (.0407) \quad (.0014) \quad (.0171) \quad (.0204)$$
$$SE=1.013,\ R2=.995,\ DW=1.56$$

2SLS estimates

$$IKF = -2.46 + .945\ IKF_{-1} - .00665\ KK_{-1} + .109\ Y - .0829\ Y_{-1}$$
$$(.6311) \quad (.0396) \quad (.00174) \quad (.0220) \quad (.02456)$$
$$SE = .8239$$

BIV estimator results

Table 7.7: Treasury bill rate

$$RS = -.151 + .892\, RS_{-1} + .01018\, GNPR - .01005\, GNPR_{-1} + .05044\, M1_{-1} + $$
$$\quad (.379)\ (.0554)\quad (.00296)\quad (.00307)\quad (.0217)$$
$$+ .142\, DD793 \cdot M1$$
$$(.0335)$$

$SE=.723$, $R2=.951$, $DW=1.62$ 2SLS estimates

$$RS = -.1426 + .913\, RS_{-1} + .00521\, GNPR - .005083\, GNPR_{-1} + .04919\, M1_{-1} +$$
$$\quad (.2996)\ (.0610)\quad (.00319)\quad (.00329)\quad (.0176)$$
$$+ .1073\, DD793 \cdot M1$$
$$(.0798)$$

$SE = .4872$ BIV estimator results

Table 7.8: Mortgage rate

$$RM = .331 + .842\, RM_{-1} + .255\, RS - .0434\, RS_{-1} - .0296\, RS_{-2}$$
$$\quad (.104)\ (.0297)\quad (.034)\quad (.0479)\quad (.0354)$$

$SE=.263$, $R2=.992$, $DW=2.14$ 2SLS estimates

$$RM = .2287 + .889\, RM_{-1} + .1880\, RS + .03772\, RS_{-1} - .09186\, RS_{-2}$$
$$\quad (.093)\ (.0247)\quad (.0533)\quad (.0952)\quad (.0581)$$

$SE = .1588$ BIV estimator results

Table 8: Comparison of Standard Error of Regression of 2SLS and BIV Estimates

equation	SE_{2SLS} (1)	SE_{BIV} (2)	% reduction SE_{BIV} vs SE_{2SLS} (3)
CS	0.746	0.597	19.9
CN	1.033	0.929	10.0
CD	1.507	1.111	26.2
IHH	1.085	0.914	15.6
Y	2.457	2.366	3.7
IKF	1.013	0.823	18.6
RS	0.723	0.487	32.6
RM	0.263	0.158	39.9

Note: Column (3) shows the reduction in the standard error of estimate (SE) of the BIV versus the (SE) of the 2SLS estimator.

Table 9: In–sample forecasts, single equation,1954Q1 –1982Q3

	2SLS		BIV		% reduction of BIV versus 2SLS	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
CN	1.049	0.807	1.046	0.803	0.28	0.49
CS	0.781	0.581	0.780	0.581	0.12	0.00
CD	1.481	1.049	1.493	1.023	–0.81	2.47
Y	3.058	2.378	3.208	2.501	–4.90	–5.17
IKF	0.991	0.748	0.997	0.743	–0.60	0.66
IHH	2.465	2.027	2.398	1.940	2.71	4.29
RS	0.824	0.580	0.841	0.567	–2.06	2.24
RM	0.258	0.189	0.260	0.178	–0.77	5.82

Table 10: In–sample model simulations,1954Q1–1982Q3

	2SLS		BIV		% reduction BIV versus 2SLS	
	RMSE (1)	MAE (2)	RMSE (3)	MAE (4)	RMSE (5)	MAE (6)
CN	1.06	0.82	1.06	0.81	0.00	1.22
CS	0.77	0.58	0.77	0.58	0.00	0.00
CD	1.50	1.06	1.51	1.03	–0.66	2.83
Y	5.82	4.69	5.96	4.80	–2.40	–2.34
IKF	1.44	1.06	1.42	1.03	1.38	2.83
IHH	2.46	2.02	2.39	1.94	2.84	3.96
RS	0.82	0.57	0.84	0.56	–2.43	1.75
RM	0.31	0.21	0.31	0.20	0.00	4.76
X	4.79	3.47	4.70	3.34	1.87	3.74
GNPR	5.82	4.69	5.96	4.80	–2.40	–2.34
V	3.06	2.37	3.21	2.49	–4.21	–5.06
KK	1.49	1.12	1.51	1.12	–1.34	0.00

Table 11: [illegible] simulations, 1982Q4–1986Q3

	2SLS		BIV		% reduction BIV versus 2SLS	
	RMSE (1)	MAE (2)	RMSE (3)	MAE (4)	RMSE (5)	MAE (6)
CN	1.27	1.16	1.25	1.14	1.57	1.72
CS	1.09	0.87	1.02	0.81	6.42	6.89
CD	3.52	3.13	3.06	2.67	13.06	14.69
Y	8.32	7.23	8.95	7.66	–7.57	–5.94
IKF	3.48	2.85	3.45	2.84	0.86	0.35
IHH	1.98	1.76	2.33	2.06	–17.67	–17.04
RS	2.16	1.89	1.79	1.56	17.13	17.46
RM	0.96	0.80	0.88	0.71	8.33	11.25
X	8.18	7.18	8.01	6.99	2.07	2.64
GNPR	8.32	7.23	8.95	7.66	–7.57	–5.94
V	8.03	6.96	8.91	7.65	–10.95	–9.91
KK	3.49	2.86	3.46	2.85	0.86	0.35

FIGURE 1: 2SLS Residuals, 75Q3-82Q3, from the RS Equation estimated with and without Oct 79 shift

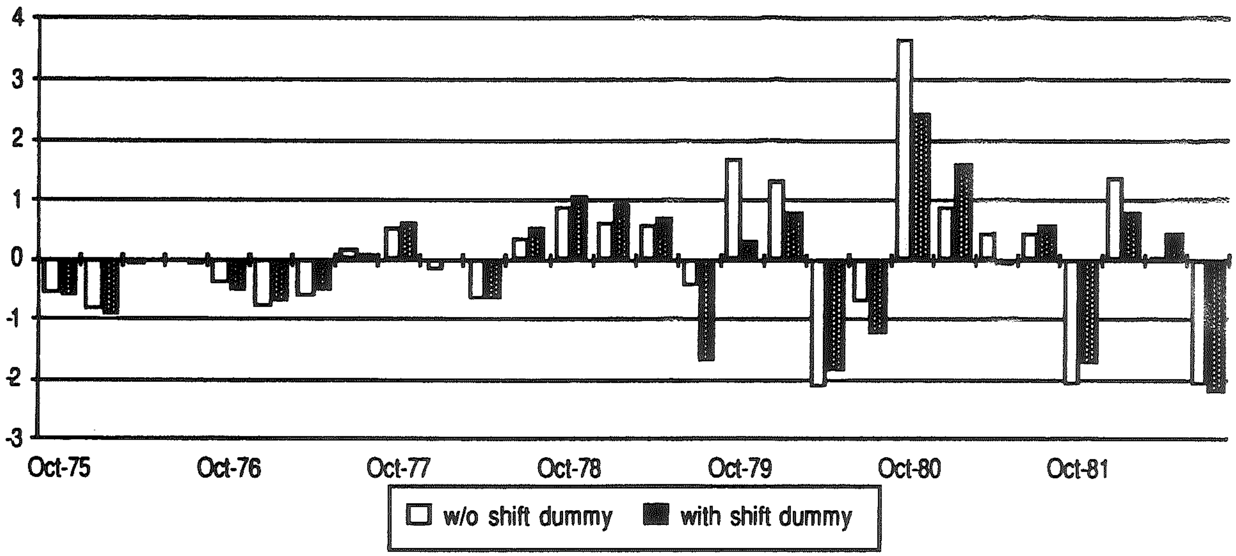

FIGURE 2: BIV Residuals, 75Q3-82Q3, from the RS Equation estimated with and without Oct 79 shift

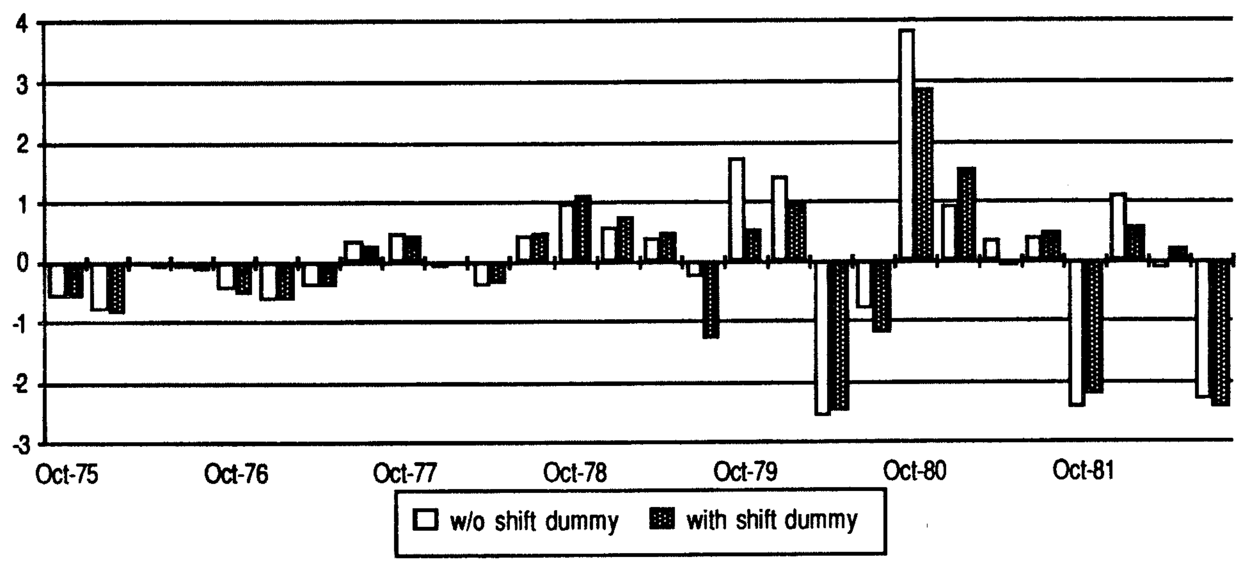

FIGURE 3: BIV Downweighting of the RS Equation, 75Q3-82Q3, estimated with and without Oct 79 shift

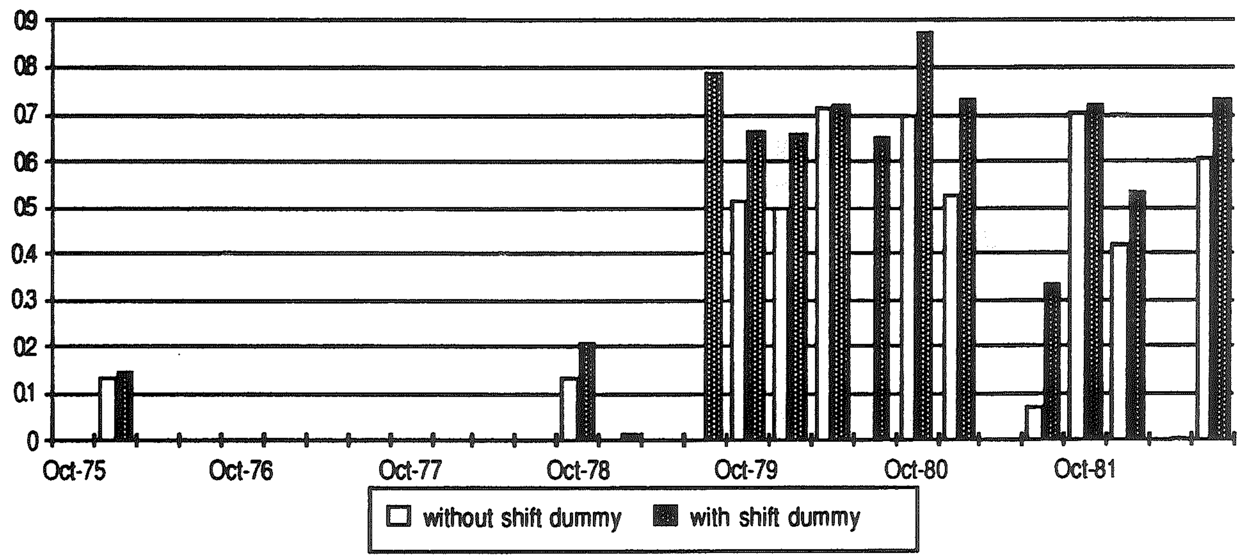

Figure 4: Residuals from Klein Model I Private Wage equation, BIV Bound varying between 2.0 and 4.0

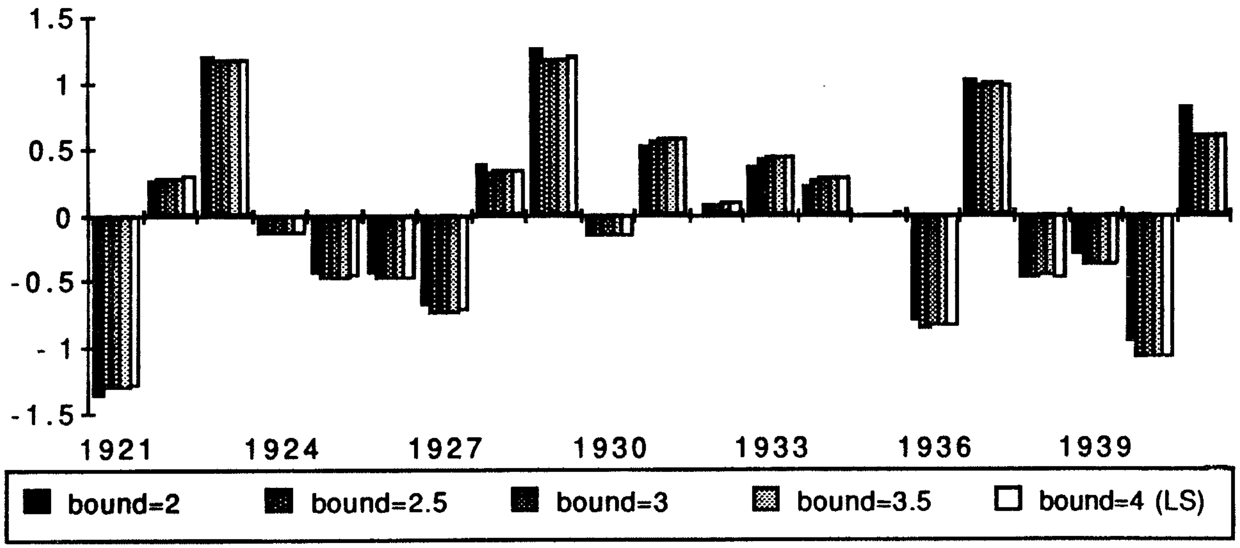

Figure 5: Estimated Coefficients from Klein Model I Private Wage Equation: BIV Bound varying between 2.0 and 4.0