ISSN 1032-3813
ISBN 0 7326 0790 6

MONASH UNIVERSITY

ANCORA · IMPARO

AUSTRALIA

THE ROBUSTNESS OF ESTIMATORS
FOR DYNAMIC PANEL DATA MODELS
TO MISSPECIFICATION

Mark N. Harris
Ritchard J. Longmire
László Mátyás

Working Paper 9/96
August 1996

DEPARTMENT OF ECONOMETRICS

The Robustness of Estimators for Dynamic Panel Data Models to Misspecification

Mark N. Harris*, Ritchard J. Longmire* and László Mátyás**

*Dept. of Econometrics, Monash University, Australia.

**Monash University and Budapest University of Economics.

Mailing address: **L. Matyas, Monash University, Department of Econometrics, Clayton 3168. Vic Australia**

Email: **Matyas@vaxc.cc.monash.edu.au**

Abstract

It is well known that the usual techniques for estimating random and fixed effects panel data models are inconsistent in the dynamic setting. As a consequence, numerous consistent estimators have been proposed in the literature. However, all such estimators rely on certain well defined assumptions, which in practice may often be violated. The purpose of this paper is to ascertain how robust the available estimators are to such misspecifications, thus providing guidance to the applied researcher as to an appropriate choice of estimator in such situations.

JEL Classification: C13, C15 and C23.

September, 1996

1. Introduction

Ever since the pioneering work of Balestra and Nerlove (1966), the estimation of dynamic error components panel data models has been at the fore of applied econometrics research.[1] The econometric interest in such models results from the fact that the usual estimation techniques (Ordinary Least Squares - OLS, Feasible Generalised Least Squares - FGLS and the Within estimators) are all biased and inconsistent in typical panel data sets *i.e.*, those with small time series components but a large cross-sectional one. This is true for both the random and fixed effects specifications (Kiviet [1995] Sevestre and Trognon [1985] and Nickell [1981]).

As a consequence, numerous semi-consistent (as N, the number of individuals, tends to infinity) estimators have been proposed in the literature, generally comprising of Instrumental Variables (IV) and Generalised Method of Moments (GMM) estimators. Without exception however, all of these estimators, either explicitly or implicitly, rely on certain assumptions or maintained hypotheses. In this paper we investigate how robust the most frequently used estimators are against some basic violations of the underlying assumptions, those which are most likely to occur in practice.

The plan of this paper is as follows. Section 2 deals with model specification, "traditional" estimation and (semi-)consistent estimators for dynamic fixed and random effects models. Section 3 describes the simulation study utilised and discusses its results. Section 4 analyses the Solow growth model previously considered in a panel context by Islam (1995) and others, and applies our previous findings to it. Finally, some conclusions are drawn in Section 5.

2. The Fixed and Random Effects Dynamic Panel Model and Traditional Estimation.

2.1 The Model

The model to be analysed is the one most commonly used in practice, that is:

$$y_{it} = \alpha_i + \delta y_{it-1} + \underline{x}'_{it}\underline{\beta} + u_{it}, \tag{1}$$

where: α_i are the *individual effects*

and: u_{it} are the usual white noise disturbance terms.

[1] Comments by Badi H. Baltagi are kindly acknowledged.

As is well known, the fixed effects specification assumes that the individual effects of (1) are fixed parameters, whereas the random effects specification treats them as random drawings from a particular distribution. It also well known that the usual methods for estimating either specification (for example, the OLS, the *Within* and Feasible Generalised Least Squares estimators) are all biased and inconsistent as $N \to \infty$ and finite T (Nickell [1981] and Sevestre and Trognon [1985]). Consistent estimators are available however, generally in the form of Instrumental Variable (IV) or Generalised Method of Moments (GMM) estimators.

Conditional on the instrument set Z, one generally has three choices of estimator for non-spherical models with variance-covariance matrix Ω (see Bowden and Turkington [1984]):

a) $\left(\tilde{X}Z(Z'\Omega Z)^{-1}Z'\tilde{X}\right)^{-1}\tilde{X}Z(Z'\Omega Z)^{-1}Z'\underline{y}$,

b) $\left(\tilde{X}\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}\tilde{X}\right)^{-1}\tilde{X}\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}\underline{y}$,

c) $\left(\tilde{X}\Omega^{-1/2}Z(Z'Z)^{-1}Z'\Omega^{-1/2}\tilde{X}\right)^{-1}\tilde{X}\Omega^{-1/2}Z(Z''Z)^{-1}Z'\Omega^{-1/2}\underline{y}$,

where $\tilde{X} = \left(\underline{y}_{-1} \vdots X\right)$. Variant a) may be recognised as the usual form of the linear GMM estimator, and is the only appropriate variant when Z contains lagged values of the endogenous variable as instruments (see Sevestre and Trognon [1995]).

2.2 Estimation in Levels.[2]

The Balestra-Nerlove Estimators: ($BN^{(L)}$) and (G-BNran).

Balestra and Nerlove (1966) show that consistent parameter estimates in an autoregressive error components (or random effects) model can be obtained by use of lagged exogenous variables as appropriate instruments (for which all three variants, a) to c), are appropriate, where Ω is the covariance matrix of the composite $v_{it} = \alpha_i + u_{it}$, disturbance term). The estimator can similarly be adapted to the fixed effects model, by estimating (1) by IV, again using lagged values of X as instruments. The $BN^{(L)}$ estimator is the only one appropriate for the fixed effects model in levels

[2] For a fuller description of the following estimators, the reader is referred to source papers and Harris and Mátyás (1996), for a useful summary.

(see below for estimators of the first differenced model). The following estimators are therefore only appropriate for the random effects specification.

The Hausman-Taylor Estimator: (HT)

Hausman and Taylor (1981) partition the X-matrix such that $X = (X_1 \vdots X_2)$, where X_1 are uncorrelated with the individual effects, but X_2 is not. In a dynamic panel data setting, the lagged dependent variable is analogous to X_2 whilst the remaining explanatory variables (X in the notation of this paper) are analogous to Hausman-Taylor's X_1. Expanding on the *G-BNran* estimator, the *HT* estimator also considers the means and deviations from the means of the original exogenous variables as valid instruments in addition to lagged values of X.

The Amemiya-MaCurdy (AM) IV Estimator

If the x's are strictly exogenous all past, present and future values become valid instruments for each time equation. Thus the Amemiya and MaCurdy (1986) estimator for the dynamic model further extends the instrument set to include X_{-1}^{*}, defined as:

$$X_{-1}^{*} = \begin{pmatrix} x_{10}^{(1)} & x_{10}^{(2)} & \dots & x_{10}^{(k)} & \dots & x_{1T}^{(1)} & \dots & x_{1T}^{(k)} \\ x_{20}^{(1)} & x_{20}^{(2)} & \dots & x_{20}^{(k)} & \dots & x_{2T}^{(1)} & \dots & x_{2T}^{(k)} \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ x_{N0}^{(1)} & x_{N0}^{(2)} & \dots & x_{N0}^{(k)} & \dots & x_{NT}^{(1)} & \dots & x_{NT}^{(k)} \end{pmatrix} \otimes \iota_T . \qquad (2).$$

This estimator, if consistent, is at least as efficient as the *HT* estimator (Amemiya and MaCurdy [1986], pp.871-872).

The Wansbeek-Bekker (WB) IV Estimator

Despite the presumed asymptotic efficiency of the GMM estimator, its small sample performance is often bettered by simpler IV type estimators. As a response Wansbeek and Bekker (1996) suggest a simple generic IV estimator. As we shall see below, the optimal *WB* estimator is that which minimises the variance of the generic one. Now both lags and leads (and linear combinations of these) of the dependent variable are included in the instrument set. That is, by defining the variable y from period $t = 1$ to $t = T$, the *WB* estimator considers linear functions of $\underline{y}_+$ as instruments, where $\underline{y}_+$ is the stacked vector of observations defined from $t = 0$ to $t = T$ for each individual. The linear functions are defined by the $(T+1)\times T$ matrix A_i, which yields $A'\underline{y}_+$ as the full instrument set (where $A = I_N \otimes A_i$). Restrictions are imposed on A_i such that:

$$A\underline{\iota}_T = 0 \qquad \text{and} \qquad \mathrm{E}\left(\underline{y}'_+ A\underline{u}\right) = \mathrm{tr}A\mathrm{E}\left(\underline{u}\underline{y}'_+\right) = 0, \tag{3}$$

where $\underline{\iota}_T$ is the $(T\times 1)$ unit vector, which respectively ensure elimination of the individual effects and consistency of the estimator.

Wansbeek and Bekker (1996) show that these conditions for A_i define its structure such that its rows sum to zero, as do each of its lowest T quasi-diagonal elements (in particular, the lower left element is zero). Wansbeek and Bekker (1996) only consider the simple AR(1) dynamic model, therefore in the case where there are additionally exogenous variables, the full *WB*-type instrument set would be defined as:

$$Z = \left(A'\underline{y}_+ \vdots X\right) \tag{4}$$

and using $\sigma_u^2(Z'Z)$ as the variance of $(Z'\underline{u})$, the generic *WB* estimator is obtained by applying GLS to equation (1) after it has been pre-multiplied by the transpose of (4).[3] The estimator's semi-asymptotic variance will be given by:

$$\sigma_u^2\left(\underset{N\to\infty}{\mathrm{plim}}\frac{1}{N}\left(\tilde{X}'P_Z\tilde{X}\right)^{-1}\right), \tag{5}$$

where: $P_Z = Z(Z'Z)^{-1}Z'$,

which, from (4), is a function of A. The optimal choice of A is that which minimises (5), such that A_i conforms to its appropriate restrictions. Harris and Mátyás (1996) suggest minimising the trace of (5) by constrained optimisation, such that A_i conforms to the restrictions of (3), treating σ_u^2 as a

[3] Note that this expression for the variance of $(Z'\underline{u})$ is only an approximation, differing from the true variance to the extent that $\mathrm{E}\left(\underline{y}'_+\underline{u}\right) \neq 0$, and this cross correlation is not taken into account.

constant. Note that the list of valid instruments can be expanded to include not only $A'\underline{y}_+$, but also $A'X_+$ for example (WB and WB^+, respectively), such that:

$$Z^+ = \left(A'\underline{y}_+ \vdots A'X_+ \vdots X\right). \tag{6}$$

These estimators can also be adapted to the model where the assumption of a scalar covariance matrix of the disturbance terms u_{it} is relaxed. The corresponding estimators are still obtained by applying GLS, but where the variance of $(Z'\underline{u})$ is now $(Z'\Omega_u Z)$, where Ω_u is unspecified and $(Z'\Omega_u Z)$ is estimated from initial preliminary estimates of $\underline{u}_i$.

The Arellano and Bover (ABov) IV Estimator

As with the WB estimator, the Arellano and Bover (1993) estimator first involves transforming the system of T equations. The nonsingular transformation is now given by:

$$H_i = \begin{bmatrix} K \\ \underline{\iota}'_T/T \end{bmatrix}, \tag{7}$$

where K is similar to Wansbeek and Bekker's A, in that $K\iota_T = 0$, where K is any $(T-1)\times T$ matrix of rank $(T-1)$. As the first $(T-1)$ transformed errors,

$$\underline{v}_i^+ = H_i\underline{v}_i = \begin{bmatrix} K\underline{v}_i \\ \bar{v}_i \end{bmatrix}, \tag{8}$$

are free of α_i, all exogenous variables are valid instruments for these first $(T-1)$ equations. Moreover, assuming serial independence of the disturbance terms v_{it}, along the lines of the Arellano-Bond estimator, the series $(y_{i0}, y_{i1}, \ldots, y_{i,t-1})$ is also a valid instrument (although this assumption requires more structure for K, which now additionally has to be upper triangular).

Letting $H = I_N \otimes H_i$, the A-Bov estimator is obtained by estimating the transformed model:

$$Z'H\underline{y} = Z'H\tilde{X}\underline{\gamma} + Z'H\underline{v}, \tag{9}$$

where: $V(Z'H\underline{v}) = Z'H\Omega_v H'Z$,

by GLS. Operationally, Arellano and Bover (*ibid* p.18), state that *provided K satisfies the above restrictions*, the *ABov* estimator is invariant to the choice of *K*.[4]

2.2 Estimation in First Differences.

As first differencing removes the individual effects, whether fixed or random, such estimators are appropriate for either specification. However, the transformed model still cannot be consistently estimated by OLS, as the lagged endogenous variable $\Delta \underline{y}_{-1}$ is correlated with the model's disturbance vector $\Delta \underline{u}$, and if the original disturbances are spherical the transformed one Δu_{it} will follow a first order moving average (MA[1]) process, with variance $\mathrm{Var}(\Delta \underline{u}) = \sigma_u^2 \Omega_\Delta$ where:

$$\Omega_\Delta = I_N \otimes \Sigma_\Delta = I_N \otimes \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}. \tag{10}$$

The Anderson-Hsiao (AH) and Arellano (AR) Estimators

Anderson and Hsiao (1982) suggest both y_{it-2} and Δy_{it-2} as an appropriate instrument for Δy_{it-1}, although the latter yields inefficient estimates and therefore y_{it-2} is a more appropriate instrument (Arellano [1988]). The small sample performance of these estimators is likely to be enhanced by including additional instruments, ΔX_{-1} for example, as if the number of instruments is the same as the number of explanatory variables, the resulting estimator will, in general, have no finite moments (Kinal [1980]).

As the "augmented" *AH* and *AR* estimators (AH^+ and AR^+) have more instruments than exogenous regressors, the non-spherical IV variant a) is now appropriate, where $\Omega = \Omega_\Delta$.[5] If one wishes to relax the assumption of a spherical original disturbance terms, following White (1984) it is possible to consistently estimate the matrix $(Z'\Omega_\Delta Z)^{-1}$ as $\left(1/N \sum Z_i' \Delta\hat{\underline{u}}_i \Delta\hat{\underline{u}}_i Z_i\right)^{-1}$. The two-step

[4] Using this result, the first difference operator was used in subsequent simulation experiments (see Section 4 below).
[5] Recalling that only variant c) is applicable when the instrument set contains lagged endogenous variables.

and one-step variants of these (and following) estimators will be asymptotically equivalent if the u_{it} are independent and homoscedastic (Arellano and Bond [1991]).[6]

The Balestra Nerlove ($BN^{(a)}$) and Sevestre-Trognon ($ST^{(b)}$ and $ST^{(c)}$) Estimators

The Balestra-Nerlove estimator can also be applied to the first difference model, where the instruments for $\Delta \underline{y}_{-1}$ are simply ΔX_{-1}. Sevestre and Trognon (1995) suggest the same instrument set, but non-spherical IV variants b) and c), as opposed to a). Although these estimators will be more efficient than those *using the same instruments* on the untransformed model (Sevestre and Trognon [1995] and White [1984]), a direct comparison with the Anderson-Hsiao estimator, for example, is not straightforward as different instrument sets are used. For the two-step estimator, Ω_Δ was directly estimated as $\hat{\Omega}_\Delta = \left(1/N \sum_{i=1}^{N} \Delta \hat{\underline{u}}_i \Delta \hat{\underline{u}}'_i \right)$.

The Arellano-Bond One Step IV (AB) Estimator

If the time series is assumed to start at $t = 0$, the variable Δy_{it-1} will only be defined at $t = 2$. At $t = 2$ the only valid instrument for Δy_{it-1} is y_{i0}. At $t = 3$, the valid set of instruments for Δy_{it-1} is now expanded to include y_{i1}, and so on. Along similar lines to the *AM* estimator, if the x's are strictly exogenous, they are also all valid instruments for each time equation (Arellano and Bond [1991]). Once the appropriate instrument set has been defined, the *AB* estimator is then a simple application of non-spherical IV variant a).

The *AB* estimator is the most semi-asymptotically efficient of all IV estimators using lagged values of the dependent variable as instruments (Sevestre and Trognon [1995]), although more efficient GMM estimators can be derived (see below). However, computationally both of the *AB* estimators may prove problematic due to: the size of the instrument matrix (especially as T increases); the loss of two time periods for estimation; and difficulty in coding the appropriate instrument matrices in standard econometric software packages.

2.3 A Generalised Method of Moments (GMM) Estimator

[6] This procedure is similarly applicable to most of the subsequent estimators, and are called "*hat*" estimators.

The essence of GMM estimation involves explicit exploitation of theoretical moment conditions which, for estimation purposes, are replaced by their sample counterparts. The IV estimators presented above are based upon only a subset (of the linear ones) of these conditions. Due to the recent work of Ahn and Schmidt (1995) and Crépon *et al.* (1996) for example, attention has turned to GMM estimation of dynamic error component panel models, although the technique can also be applied to fixed effects dynamic models.

Firstly define the initial values as:

$$y_{i0} = \alpha_i + \underline{x}'_{i0}\underline{\beta} + u_{i0}, \tag{11}$$

where the parameter vector corresponding to the exogenous variables β is assumed to be identical across equations (1) and (11) (a requirement of the need for consistent starting values of the full parameter vector for GMM estimation). Equations (1) and (11), along with the "usual" set of standard assumptions, allow q implicit orthogonality conditions to be defined. These conditions for both the fixed and random effects specifications are summarised in Table 1 below, along with how these relate to the IV estimators presented above.

Table 1: GMM Conditions and IV Estimators[1]

	Fixed Effects Model		Random Effects Model	
	Condition	Estimators using it	Condition	Estimators using it
a.	$\mathrm{E}(u_{i0})=0$	GMM	$\mathrm{E}(v_{i0})=0$	GMM
b.	$\mathrm{E}(u_{i0})^2=\sigma_0^2$	GMM	$\mathrm{E}(v_{i0})^2=\sigma_0^2$	GMM
c.	$\mathrm{E}(u_{i0}u_{it})=0$	All except *AH*, *AH*$^+$ & 2-step *ST*	$\mathrm{E}(v_{i0}v_{it})=\sigma_\alpha^2$	All except *AH*, *AH*$^+$ & 2-step *ST*
d.	$\mathrm{E}(u_{it})=0$	GMM	$\mathrm{E}(v_{it})=0$	GMM
e.	$\mathrm{E}(u_{it}u_{is})=0$	All except 2-step $BN^{(L)}$, $BN^{(\Delta)}$ & $ST^{(b)}$	$\mathrm{E}(v_{it}v_{is})=\sigma_\alpha^2$	All except 2-step $BN^{(L)}$, $BN^{(\Delta)}$ & $ST^{(b)}$
f.	$\mathrm{E}(u_{it}^2)=\sigma_u^2$	All one-step estimators	$\mathrm{E}(v_{it})^2=\sigma_u^2+\sigma_\alpha^2$	All levels estimators except GMM, *WB*, *WB*$^+$ & *ABov*
g.	$\mathrm{E}(u_{i0}x_{it}^k)=0$	GMM	$\mathrm{E}(y_{i0}v_{it})=c^2$	GMM, *AR*, *AR*$^+$, *AB* & *AB*$^+$
h..	$\mathrm{E}(u_{it}x_{i0}^k)=0$	All except *AR*, *AH* & *AB*	$\mathrm{E}(v_{i0}x_{it}^k)=0$	GMM
i.	$\mathrm{E}(u_{it}x_{it}^k)=0$	All	$\mathrm{E}(v_{it}x_{i0}^k)=0$	All except *AR*, *AH*, *AB* & *WB*
j.	$\mathrm{E}(u_{it}x_{is}^k)=0$	All except *AR*, *AH* & *AB*	$\mathrm{E}(v_{it}x_{it}^k)=0$	All
k.			$\mathrm{E}(v_{it}x_{is}^k)=0$	All except *AR*, *AH*, *AB* & *WB*

Notes:
[1] The relationship between conditions and IV estimator is only an approximation in many cases, for example estimators that use X_{-1} as instruments obviously do not require independence between *all* of the disturbances and *all* of the X's, *etc.*.
[2] c is a constant.

Harris and Mátyás (1996) give an example of how these orthogonality conditions translate into identifying equations expressed in terms of observed variables and parameters. The GMM estimator of the full parameter vector $\underline{\psi}$, which contains the parameters of interest as well as other nuisance parameters, is simply obtained by the value that minimises the criterion function:

$$\underline{\hat{\psi}} = \min_{\psi} m_N(\underline{\psi})' \hat{W}^{-1} m_N(\underline{\psi}) \tag{12}$$

where: $m_N = N^{-1}\sum_i m_i(\underline{\psi})$,

$$W = \lim_{N\to\infty} \operatorname{cov}\left(N^{-1/2}\sum_i m_i\right) = \operatorname{cov}(m_N)$$

and: $\hat{W} = N^{-1}\sum_i m_i(\underline{\psi})\, m_i(\underline{\psi})'$, evaluated at an initial consistent estimate of $\underline{\psi}$.

A problem with such GMM estimators is that the number of such orthogonality conditions q tends to infinity as $T \to \infty$. The question of how many of these over-identifying equations one should use therefore naturally arises. Although asymptotic efficiency arguments suggest all such conditions should be used, Crépon *et al.* (1996) have shown that there is no efficiency loss in disregarding those equations in which any of the parameters of interest do not feature (which unfortunately does not apply in this case). Moreover, Keonker and Machado (1996) show that if q increases without bound as the sample size (here in T) increases relative to the dimensions of the parameter vector, the usual limiting distribution of the GMM estimator may not be valid. The problem arises in the estimation of $\hat{W}$ as $q \to \infty$ too quickly. An obvious example is that the columns of $\hat{W}$ are likely to become increasingly collinear, such that it becomes singular. Thus, as in Harris and Mátyás (1996) and after Keonker and Machado (1996), two GMM-type estimators are considered. The first one uses the (numerically determined) maximum number of such conditions, with the "optimal" weighting matrix, $\hat{W}$. The second uses all such conditions, but uses I as the weighting matrix (Table 2).

Table 2: GMM-Type Estimators for the Dynamic Panel Model.

Sample Size		Conditions Used			
T	N	Fixed Effects	Random Effects	Weighting Matrix[1]	Estimator Mnemonic
4	25	a) - j)	a) - k)	I	GMM_F1(R1)
10	25	a) - j)	a) - k)	I	GMM_F1(R1)
4	25	a) - g) and i)[2]	a) - g), j)	$\hat{W}$	GMM_F2(R2)
10	25	a), b), d) and f)	a), b), d) and e)	$\hat{W}$	GMM_F4(R4)

Notes: [1] $\hat{W}$ is the estimated covariance matrix of the empirical moments. [2] 2i) for $k = 1$ only.

3. The Simulation Experiments

The data for the simulation experiments follows closely that Harris and Mátyás (1996), such that the *assumed* data generating process was:

$$\begin{aligned}
&y_{it} = \alpha_i + \delta y_{it-1} + x_{it}^{(1)}\beta_1 + x_{it}^{(2)}\beta_2 + u_{it}, \\
&y_{i0} = \alpha_i + \underline{x}'_{i0}\underline{\beta} + u_{i0} \\
&\text{where: } u_{it} \sim iid\,\mathrm{N}(0,1), \\
&\qquad x_{it}^{(k)} = \rho_x x_{i,t-1}^{(k)} + \text{uniform}(-0.5, 0.5),\ k = 1,2 \\
&\quad \text{and: } \delta = \beta_1 = \beta_2 = 0.5,\ \rho_x = 0.5 \text{ and } 1.
\end{aligned} \tag{13}$$

The individual effects were generated as $\alpha_i = 1,\ldots,N$ and $\alpha_i \sim iid\,\mathrm{N}(0,1)$ for the fixed and random effects specifications respectively. Sample sizes of $T = 4$, 10 and $N = 25$, 50 were chosen. Finally, due to computation time, the number of Monte Carlo repetitions was limited to 100.[7] In each case, analysis is focussed upon the estimation of δ.

3.1 Misspecification: The Scenarios

The violations of the usual assumptions for the fixed effects model considered were that the disturbances were allowed to be autocorrelated, such that:

$$u_{it} = \rho u_{i,t-1} + \zeta_{it}, \qquad \zeta_{it} \sim iid\,\mathrm{N}(0,1). \tag{14}$$

Secondly, the "exogenous" variables were allowed to be correlated with the disturbance terms:

$$x_{it}^{*} = x_{it} + \theta u_{it}^{*}, \qquad u_{it}^{*} \sim iid\ \mathrm{N}(0,1), \tag{15}$$

[7] For example, the sample size of $T = 10$ and $N = 50$ was not undertaken for the random effects model as it was estimated that the simulation would take in excess of two months on a Pentium 120 personal computer.

For the random effects specification, violations of the usual assumptions were again that the disturbances were allowed to be serially correlated, *as per* equation (14). A correlation was also instigated between the individual effects and the explanatory variables:

$$x_{it}^{*} = x_{it} + \gamma\alpha_i , \tag{16}$$

and finally, the individual effects and the disturbances were correlated:

$$u_{it}^{*} = u_{it} + \gamma^{*}\alpha_i . \tag{17}$$

The strength of all of these correlations varied from weak (with the parameters ρ, γ, θ and γ* equal to 0.1) to medium (0.5) to strong (0.9). For the smallest sample size considered ($T = 4$, $N = 25$), a fuller range of parameters was considered (0.01 - 0.99).

These sources of misspecification were chosen not only because they were thought to be the most likely to occur in applied work, but also because they should directly affect certain estimators, and to varying extents (see Table 1). For example, serially correlated errors invalidates the past history of y as an instrument, as well as making estimators less efficient. If the x variables are correlated with disturbances, those IV estimators using lagged x's as instruments should fare poorly (although this should be less prevalent in the random effects model in first differences). Correlating the individual effects and the disturbances should adversely affect those estimators that assume independence of such (primarily those operating in levels in which the individual effects are not removed, and that construct a composite covariance matrix). Finally, the GMM-type estimators should be adversely affected by all of these sources of misspecification as they explicitly rely on all of them.

In analysing the results for the experiments, we focus on Mean Squared Errors (MSE's) of the estimation of δ in jointly assessing the estimators' performance both in terms of bias and variance (the results β and for the absolute biases and ranges of such can be obtained from the authors on request). Moreover, although the results presented below are for $\rho_x = 1$, the only substantive differences in setting $\rho_x = 0.5$ were that though absolute MSE's were different, the rankings across estimators remained substantially unaltered. A selection of the results is presented in Figures 1 to 8 in the Appendix.

3.2 The Simulation Result: Random Effects Model

In both samples and for both the levels and the Δ model, several points immediately arise. Firstly, some estimators can be immediately disregarded in terms of excessive MSE. For example, those estimators that use expanded instrument sets for each time equation (*ABov* and AB^+) tend to suffer heavily from both the resulting small sample bias (*i.e.*, when the correlation parameters equal zero) and the misspecification bias. Those estimators that have no finite moments (the simple *AR* and *AH*), also unsurprisingly, tend to have poor have small sample performance and misspecification bias. Several estimators exihibited almost identical performance, for example all variants of the *AM* estimator, the FGLS and OLS estimators (in levels), and invariably the two-step estimators when *T* is "moderate" ($T = 10$).

AR Residuals

In the levels model, in addition to those poorly performing estimators as outlined above, the $BN^{(b)}$ $BN^{(c)}$ and estimators can be disregarded in the small *T* sample (Figure 1), although their performance improves with *T*. Many of the levels estimators follow a similar "cyclical' pattern, especially discernible in the smallest sample size (*BN* and *HT* variants). Only the GMM-type, *WB*, *AM* and (inconsistent) OLS, *Within* and FGLS levels estimators exhibit a stable performance across likely values of ρ. The performance of these does though vary with ρ. The *WB* estimators in the small sample improve significantly with ρ (with the augmented instrument set variant WB^+ outperforming its simpler counterpart, although there is very little difference between the one and two step estimators), as does that of the *Within* (again, at least in the small sample). As one might expect, the performance of the OLS and FGLS estimators are adversely affected by increasing ρ (as was the *AM* estimator). Although this is generally the case for the GMM-type estimators, the effect is much reduced, again especially in the small sample. In the larger *T* sample, the *Within*, *WB* and WB^+ estimators yield both low and stable MSE's, with the latter two outperforming the former apart from mid range values of ρ (≈ 0.5).

As ρ tends to unity, the covariance matrix of the first differenced disturbances does tend to equation (10), differing only by a scale factor. Indeed, in the small sample most of the estimators' performance improves with ρ. However, the consistent estimators are bettered by OLS at every value (Figure 2) and the $ST^{(a)}$, $ST^{(b)}$ and $BN\ \Delta$ estimators can be disregarded, in addition to those previously mentioned, due to excessive MSE's. Increasing T does however, improve the relative performance of the consistent estimators, most notably the *AB* estimator, which is quite clearly the dominant first difference estimator in samples of moderate T.

In summary, when the disturbances are serially correlated and T is small, several (consistent) first difference estimators have good performances, but their MSE functions are erratic in ρ. Therefore an appropriate estimator would appear to be a choice between the consistent GMM or *WB* estimators, or the computationally easier, but inconsistent, *Within* and OLS Δ estimators. In moderate (or large) T the *AB* estimator performs consistently well across ρ

Individual Effects and Disturbances Correlated

In the levels model for the smallest sample size, the GMM estimators uniformly dominate at all levels of correlation between the α_i and the u_{it} (Figure 3). Presumably due to the misspecification in the assumed covariance matrix, most of the other estimators appear to be adversely, and unpredictably, affected by such a correlation, for example now the $BN^{(a)}$, $HT^{(a)}$ and $HT^{(b)}$ can also be immediately disregarded. The exceptions are the *AM*, *Within*, OLS, FGLS and *WB* estimators. The MSE's of all but the latter increase monotonically with the correlation $\gamma*$, whereas the *WB* estimators' performance actually improves with increased correlation. The GMM estimators again perform well when T is moderate. However, at all levels of correlation, all four variants of the *WB* estimator uniformly dominate the GMM estimators, as does the simple *Within* estimator, although the clearly dominant estimator is the WB^{+} one.

Of the first difference estimators in small T samples, many of their MSE functions are both large and erratic in $\gamma*$ (Figure 4). For a predictable low MSE one might be tempted to use the OLS Δ estimator for low values of correlation between the individual effects and the disturbances, or any of the AR^{+}, AR^{+} *hat* and *AB* estimators for medium to strong correlation. With moderate T, the *AB* estimator clearly dominates the remaining Δ estimators, having the lowest (and stable) MSE (although the $ST^{(b)}$ estimator also fares rather well). However, given the smaller MSE's found in the levels estimators, the preferred estimators would be GMM and WB^{+} for small and moderate T respectively.

Individual Effects and X Correlated

Violating the exogeneity assumption of X severely affects most of the estimators operating in levels and using X in some form as an instrument in the small T samples (Figure 5). At weak levels of correlation, the GMM estimators again dominate the others. However, somewhat surprisingly, from $\theta = 0.2$ and stronger, the simple OLS and FGLS estimators clearly dominate in the small samples. In moderate T samples, the simple FGLS and OLS estimators again fare well, especially at stronger levels of correlation. Again performing well at low levels of correlation, the GMM estimators are more adversely affected at strong levels of correlation in this larger sample, and would not be recommended. The *Within* estimator's reasonable and stable performance appears even more attractive in moderate T. However, for a consistent estimator, the WB^{+} and WB^{+} *hat* estimators exhibit a low and robust MSE in the larger sample sizes, and indeed have quite reasonable performance in small T (although the simpler WB estimators fare rather poorly in both samples).

With small T none of the Δ estimators clearly dominates the others (Figure 6), although the OLS estimator has a stable and reasonable performance across θ. Overall, although again the simple AR and AH variants are not appropriate, their augmented instrument set counterparts could be considered, as could the ST variants. Surprisingly, of all the AB variants, only the two-step AB^{+} estimator has reasonable performance. These results were substantially unaltered when T was increased to 10, although now the two-step AB^{+} estimator can be disregarded. For such misspecification therefore, the WB^{+} again appears to be a good choice.

3.3 The Simulation Result: Fixed Effects Model

Results consistent across all simulations are firstly that expanding the instruments sets of the AR and AH estimators, and indeed allowing for an unspecified covariance matrix, does not improve their performance. Surprisingly, the performance of these estimators is by far superior in the fixed effects specification. Once more the AB^{+} estimator appears to suffer especially from both small sample and misspecification bias in N, when T is small or moderate. Finally, the $ST^{(b)}$ and BN estimator in levels are effectively identical, as are the one and two step AB estimators, even in small T samples.

AR Residuals Results

Again many of the estimators follow a "cyclical" pattern with ρ, with a distinct peak at $\rho = 0.5$ (Figure 7). In both samples several estimators (GMM, *AH*, *AR*, *AB*, *Within* and OLS Δ) clearly dominate, having both a very small and stable MSE, of which the latter two may be favoured, again in terms of ease of computation.

Disturbances and X Correlated

In the smallest sample size, as expected, most of the estimators performance decreases with γ. However, any of the *Within*, OLS Δ, *AB*, *AH*, *AR* and GMM estimators, provide a small and relatively constant MSE against all values of γ (Figure 8), of which the simpler *Within* and OLS Δ may again be preferred.

We also carried out a "Semi" Monte Carlo experiment, where the exogenous variables were not generated as in (13), but instead taken from Section 4 and kept fixed. Then formulae (14) - (17) were used to generate u_{it}, u_{it} * and x_{it}*. Although the absolute values of the biases and the MSE's obtained were different form those obtained in the "true" Monte Carlo experiments, the basic pattern of behaviour of the different estimator was not affected by this. The poorly performing estimators remained poor and the recommended estimator continued to perform well. This shows our findings are not limited to the setup of our experiments, and can genuinely be used as guidelines for empirical applications.

4. An Empirical Application

Let us now see how the above results can be used in the case of a "real data" application. The Solow model (Solow [1956]) seems to be a good candidate for this exercise, as it has an important role in neoclassical theory and has recently been estimated on panel data sets in several studies (see, for example, Islam [1995], Nerlove [1996], Smith, Lee and Pesaran [1996]). It assumes (in the form used here) a constant return-to-scale production function, substitution between the two inputs labour and capital and constant depreciation and technical change. Convergence of the different countries' growth rates (starting from different GDP per capita levels) is then deduced from the theory. The model used in the above studies has the form

$$y_{it} = \delta y_{it-1} + x_{it}'\beta + \alpha_i + u_{it}$$

where y_{it} = ln(Real GDP per capita, measured in Terms of Trade prices)

x_{it} = [ln(Real Inv), ln(Population growth rate + tech change + depreciation rate)].

We estimated the model for 22 OECD countries for the period 1950-1990 using annual data downloaded from the Penn World Tables 5.6. The sample size considered for the empirical application is N=22 and T=41. Some of the estimators such as the Arellano-Bond and Arellano-Bover estimators are numerically difficult to estimate as T increases since the number of columns in the matrix of instruments increases substantially, causing serious multicollinearity problems. The WB estimator could not be estimated using T=41 since the objective function would have to be minimised over more than 1000 parameters. The sample was also transformed to quinquennium data using the levels at the end of each five year period so that T=9, a number comparable to the simulation experiments.

Table 3 Differenced Model Estimators.*

		5-yearly data (T=9)	***Speed of Convergence***		***Annual Data (T=41)***	***Speed of Convergence***
	δ	***S.E(δ)***	***% per year***	δ	***S.E(δ)***	***% per year***
OLS	*0.773*	0.034	5.16	*0.544*	0.024	60.91
GLS	*0.869*	0.014	2.81	*0.967*	0.003	3.40
AR	*0.909*	0.017	1.91	*0.964*	0.007	3.67
AR+	*0.843*	0.016	3.42	*0.945*	0.007	5.71
AR+ hat	*0.831*	0.001	3.70	*0.909*	0.0003	9.53
STa	*0.769*	0.025	5.24	*0.865*	0.009	14.52
STb	*0.747*	0.025	5.83	*0.705*	0.012	34.96
STa hat	*0.669*	0.004	8.05	N/A	N/A	N/A
STb hat	*0.628*	0.004	9.30	N/A	N/A	N/A
AB	*0.864*	0.014	2.93	N/A	N/A	N/A
AB hat	*0.873*	0.001	2.71	N/A	N/A	N/A
AB+	*0.827*	0.012	3.80	N/A	N/A	N/A
BN	*0.506*	0.101	13.63	*0.205*	0.056	158.26
BN hat	*0.473*	0.008	14.97	*0.206*	0.002	157.93
AH	*0.905*	0.024	2.00	*0.841*	0.028	17.37
AH+	*0.881*	0.056	2.53	*0.688*	0.139	37.37
AH+ hat	*0.869*	0.005	2.81	*0.575*	0.008	55.37

* Technical change parameter, exogenously set at 0.05 (per five year period). Depreciation parameter, exogenously set at 0.2 (per five year period). OECD Countries (N=22, T=9 and 41)

Table 4 Levels Model Estimators [*]**.**

		5-yearly data (T=9)	*Speed of Convergence*		*Annual Data (T=41)*	*Speed of Convergence*
	δ	*S.E*(δ)	*% per year*	δ	*S.E*(δ)	*% per year*
OLS	*0.904*	0.009	2.01	*0.976*	0.002	2.47
FGLS	*0.896*	0.010	2.20	*0.973*	0.002	2.77
Within	*0.869*	0.012	2.81	*0.964*	0.003	3.64
GBNran 2a	*0.812*	0.027	4.18	*0.897*	0.010	10.90
GBNran 2b	*0.799*	0.026	4.50	*0.902*	0.010	10.30
GBNran 2c	*0.765*	0.032	5.37	*0.899*	0.011	10.64
HT 2a	*0.828*	0.023	3.77	*0.918*	0.009	8.50
HT 2b	*0.827*	0.022	3.80	*0.908*	0.009	9.60
HT 2c	*0.815*	0.026	4.08	*0.909*	0.011	9.57
AM 2a	*0.892*	0.012	2.29	*0.968*	0.003	3.21
AM 2b	*0.892*	0.012	2.29	*0.969*	0.003	3.13
AM 2c	*0.895*	0.012	2.21	*0.970*	0.003	3.06
WB	*0.868*	0.019	2.83	N/A	N/A	N/A
WB+	*0.876*	0.015	2.64	N/A	N/A	N/A
WB hat	*0.872*	0.017	2.73	N/A	N/A	N/A
WB+ hat	*0.881*	0.014	2.54	N/A	N/A	N/A
GMM(I)	*0.105*	0.012	45.08	N/A	N/A	N/A
GMM(W)[+]	*0.274*	0.044	25.89	N/A	N/A	N/A

[*] See notes to Table 3. [+] Uses only conditions a and d (Table 1).

The speed of convergence λ is related to δ as $\delta = \exp(-\lambda\tau)$, so that real GDP per capita will converge to the steady state if $0 < \delta < 1$.

The very first impression from the estimation result presented in Tables 3 and 4 is that the estimated δ parameter values and the implied speeds of convergence are quite different between estimators for both data sets, but also between the two time series used. The estimated parameters are uniformly larger for the annual data set with several highly unlikely implied convergence rates for both time series. Some of the results for the quinquennium data are similar to those presented by Islam (1995) and Nerlove (1996) but many are markedly different. Note also, that the estimators which do not contain the dependent variable in the instrument set (ST and BN) produce the lowest estimates of δ. Given the relatively large *T*, the differences in the estimates are due to the efficiency and/or small sample performance of the estimators (Harris and Mátyás [1996]). The most plausible explanation may be that some of the basic assumptions behind the utilised estimators are violated, and thus the produced parameters values are unreliable.

The simulation study showed that for the larger sample size, the *WB* and *Within* estimators are the most robust against the three types of misspecification considered. For the quinquennium data the derived convergence rates are nearly identical and make good economic sense. The bottom line here is that one must be very careful about the choice of the appropriate estimator, and should not automatically assume that the underlying assumptions behind a model are satisfied. This implies that robustness must have an important role when choosing an estimator for empirical applications.

5. Conclusions

In this paper we considered several well known estimators for the dynamic panel model (for example, *AR*, *AH* and *AB*) as well as other less well known ones (GMM and *WB* for fixed and random effects model respectively). All of these estimators rely on certain assumptions which we purposely violated to ascertain how robust they are to such misspecification. In this way the applied researcher can be confident in his/her choice of estimator, even if the assumed data generating process is not true.

Two-step "robust" variants of many of these estimators were also considered. Invariably the two step variants showed no significant improvement on their (misspecified) one step counterparts. This (superficially) surprising result was presumably a consequence of the fact that the initial estimates of the disturbance term were, dependent upon the exact form of misspecification, invariably inconsistent, as would be any estimate of their covariance matrix based upon these. For example, AR disturbances invalidate the recent past history of $y_{i,t-1}$ as an instrument. Accordingly, estimators such as the *AR* one will be inconsistent, as will any estimates of disturbances (and covariances of such) based upon these. These findings bring into question the usefulness of two-step estimators such as the *AB*, *AR*, *AH* and, indeed, the GMM ones.

In terms of preferred random effects estimators, if one is concerned about the true data generating process, in small T samples the GMM-type estimators generally have a reasonable performance against most types and strengths of misspecification. In moderate T samples the WB^{+} estimator tended to be the dominant one. These estimators may prove computationally burdensome, and some researchers may prefer either of the (inconsistent in the case of no-misspecification) *Within* or OLS Δ estimators. The choice of estimator appears less important for the fixed effects model. Several estimators had a similar (and acceptable) performance, most notably the GMM, *AR*, *AH*, *AB*, OLS Δ and *Within* estimators (the latter two being inconsistent in the "usual" setting). Although unlike in the random effects setting, where prudence suggests using the expanded instrument sets of the *AR* and *AH* estimators (if indeed these estimators are chosen), in the fixed effects model there appears to be very little difference between the three variants of these particular estimators. However, for ease of computation the OLS Δ and *Within* estimators appear an acceptable choice.

References

Ahn, S.C. and Schmidt, P. (1995); Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, Vol. 68, pp. 5-28.

Anderson, T.W. and Hsiao, C. (1982); Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics*, Vol. 18, pp. 578-606.

Amemiya, T. and MaCurdy, T.E. (1986); Instrumental Estimation of an Error Components Model, *Econometrica*, Vol. 54, pp. 869-881.

Arellano, M. (1988); A Note on the Anderson-Hsiao Estimator for Panel Data, *mimeo* Institute of Economics, Oxford University.

Arellano, M. and Bond, S. (1991); Some Tests of Specification for Panel Data: Monte-Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, Vol. 58, pp. 127-134.

Arellano, M. and Bover, O. (1993); Another Look at the Instrumental Variables Estimation of Error-Components Models, *Journal of Econometrics*, Vol. 68, pp. 29-52.

Balestra, P. and Nerlove, M. (1966); Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model, *Econometrica*, Vol. 34, pp. 585-612.

Breusch, T.S., Mizon, G.E. and Schmidt, P. (1989); Efficient Estimation Using Panel Data, *Econometrica*, Vol. 57, pp. 695-700.

Crépon, B., Kramarz, F. and Trognon, A. (1996); Parameters of Interest, Nuisance Parameter and Orthogonality Conditions: An Application to Error Component Models, *forthcoming Journal of Econometrics*.

Harris, M.N and Mátyás, L. (1996); A Comparative Analysis of Different Estimators for Dynamic Panel Data Models, Working Paper 4/96, Monash University, Melbourne, Australia.

Hausman, J.A. and Taylor, W.E. (1981); Panel Data and Unobservable Individual Effects, *Econometrica*, Vol. 49, pp. 1377-1398.

Islam, N. (1995); Growth Empirics: A Panel Data Approach, *Quarterly Journal of Economics*, Vol. 110, pp. 1127-1170.

Kinal, T.W. (1980); The Existence of *k*-class Estimators, *Econometrica*, Vol. 48, pp. 241-249.

Kiviet, J.F (1995); On Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Data Models, *Journal of Econometrics*, Vol. 68, pp. 53-78.

Keonker, R., and Machado, J.A.F. (1996); GMM Conference When the Number of Moment Conditions is Large, mimeo.

Nerlove, M. (1996); Growth Rate Convergence, Fact or Artefact?, paper presented at the Sixth Biennial International Conference on Panel Data, Amsterdam, June 1996.

Nickell, S. (1981); Biases in Models With Fixed Effects, *Econometrica*, Vol. 49, pp. 1417-1426.

Sevestre, P. and Trognon, A. (1985); A Note on Autoregressive Error Component Models, *Journal of Econometrics*, Vol. 29, pp. 231-245.

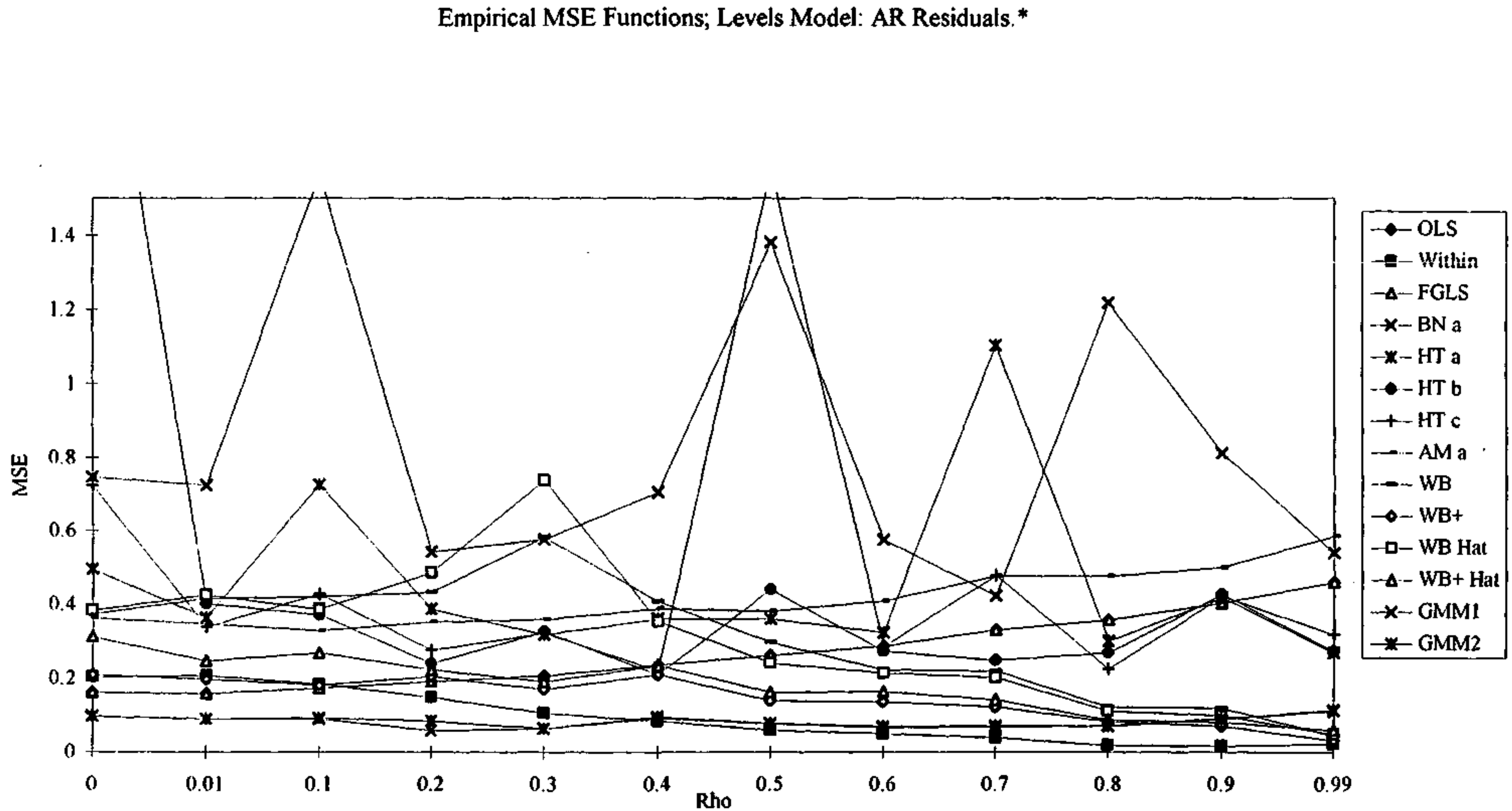
Sevestre, P., and Trognon, A. (1996); Dynamic Linear Models, in (chpt 7) *The Econometrics of Panel Data*, second revised edition, Mátyás and Sevestre (eds.), 1996, Kluwer Academic Publishers, Dordrecht.

Smith, R., Lee, K. And Pesaran, H. (1996); Growth and Convergence: A Multi-country Empirical Analysis of the Solow Growth Model, paper presented at the Sixth Biennial International Conference on Panel Data, Amsterdam, June 1996.

Wansbeek, T. and Bekker, P. (1996); On IV, GMM and ML in a Dynamic Panel Data Model, *Economic Letters*, Vol. 51, pp. 145-152.

White, H. (1984); *Asymptotic Theory for Econometricians*, Academic Press, Orlando.

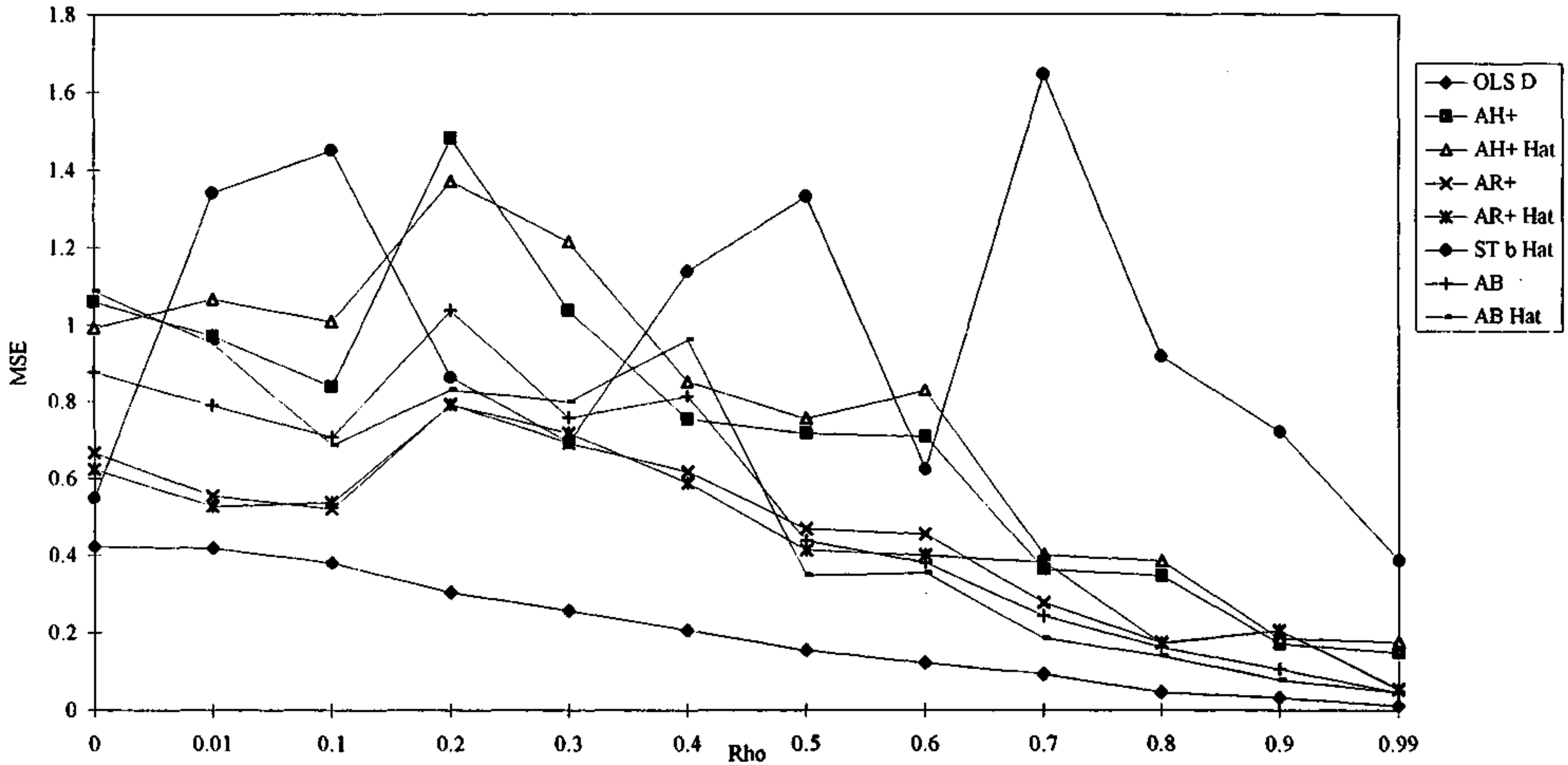
Figure 1: Random Effects Model; N=25, T=4



**ABov*, *ABov hat* and *BN (b and c)* excluded due to excessive MSE. *AM (b and c)* excluded as identical performance to *AM (a)*.

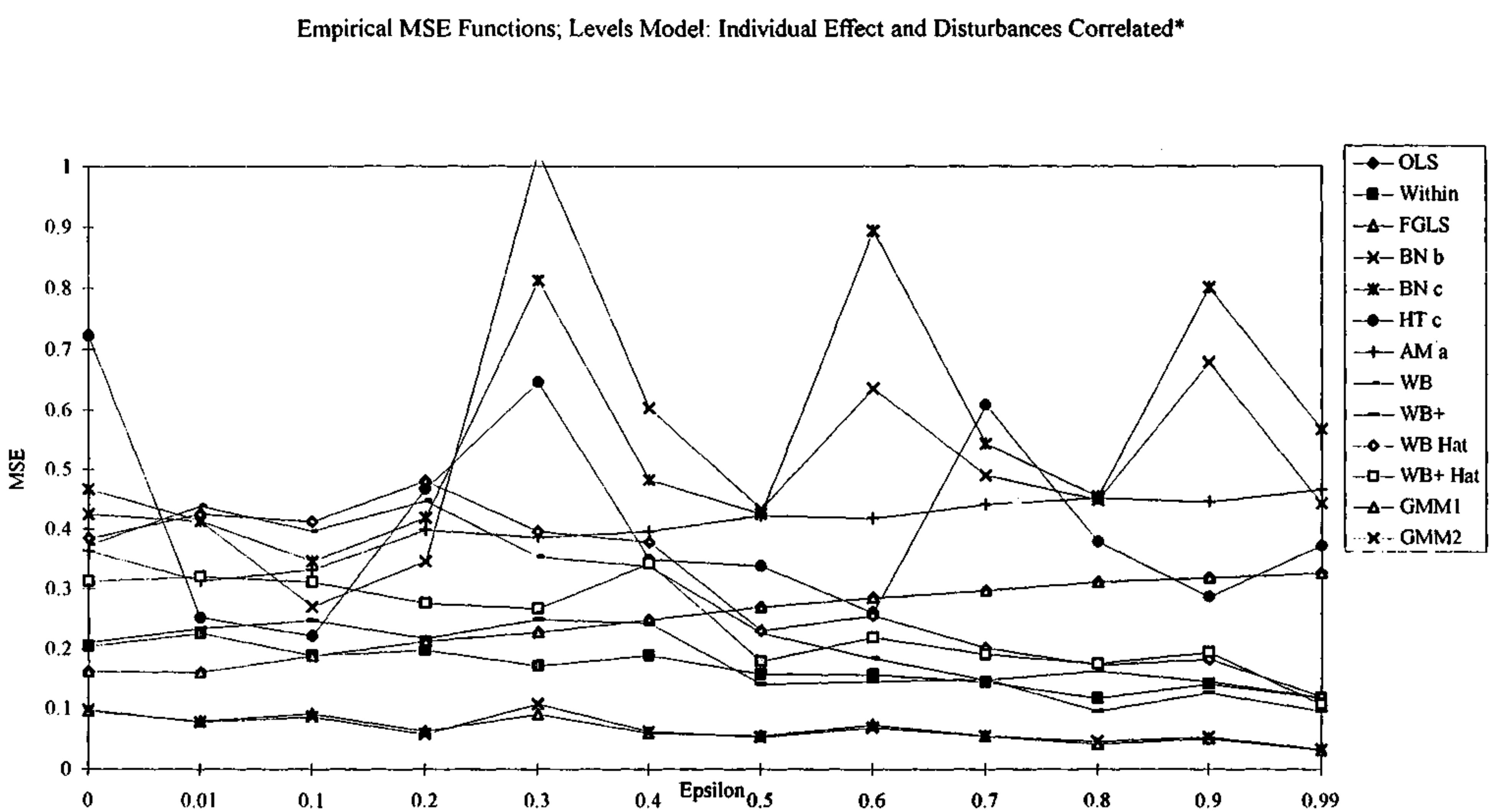
Figure 2: Random Effects Model; N=25, T=4

Empirical MSE Functions; Δ Model: AR Residuals.*



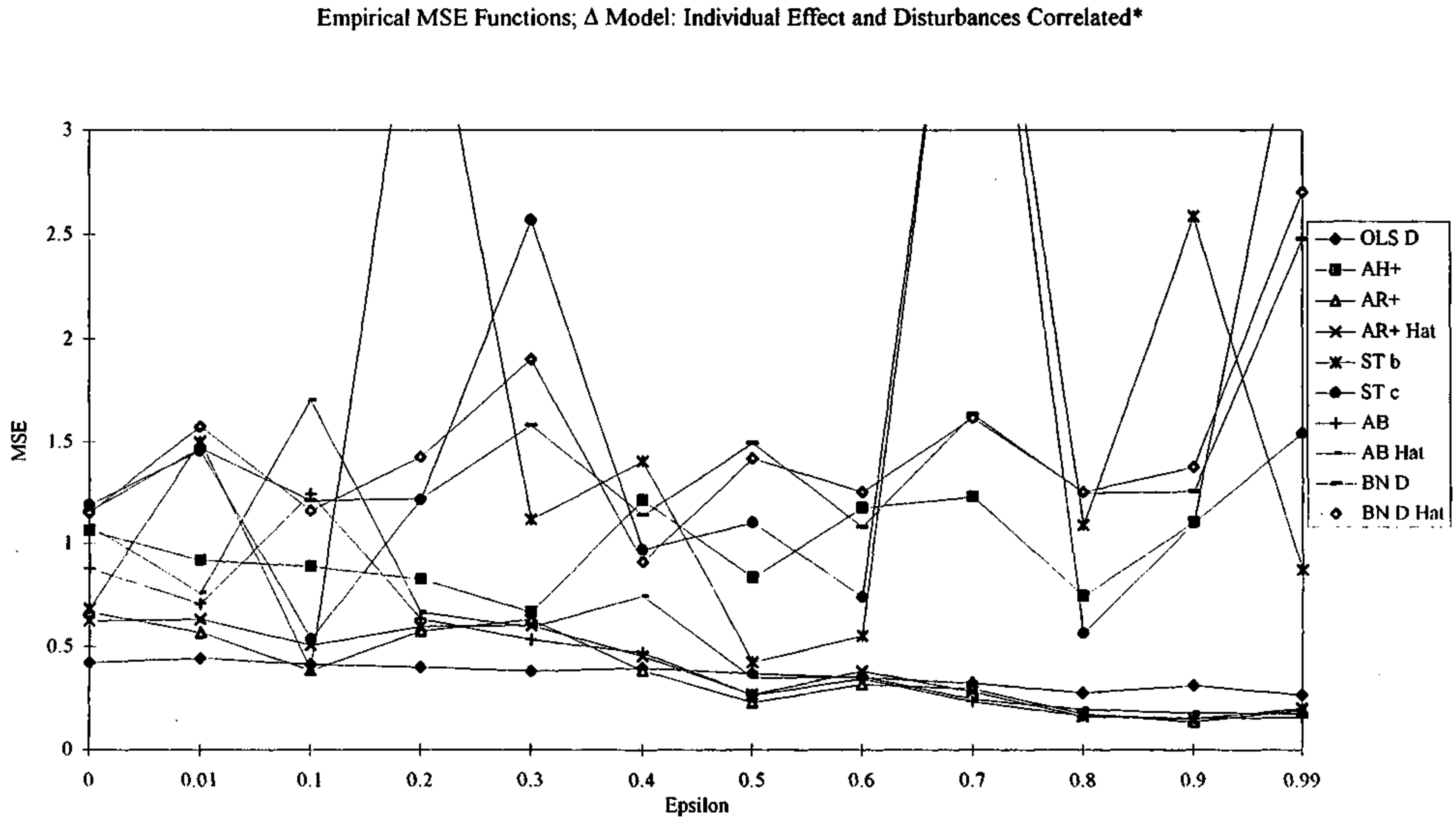
**AR*, *AH*, *AB+*, *AB+ Hat*, *ST (a and b)*, *BN* Δ and *BN* Δ *Hat* not included due to excessive bias.

Figure 3: Random Effects Model; N=25, T=4



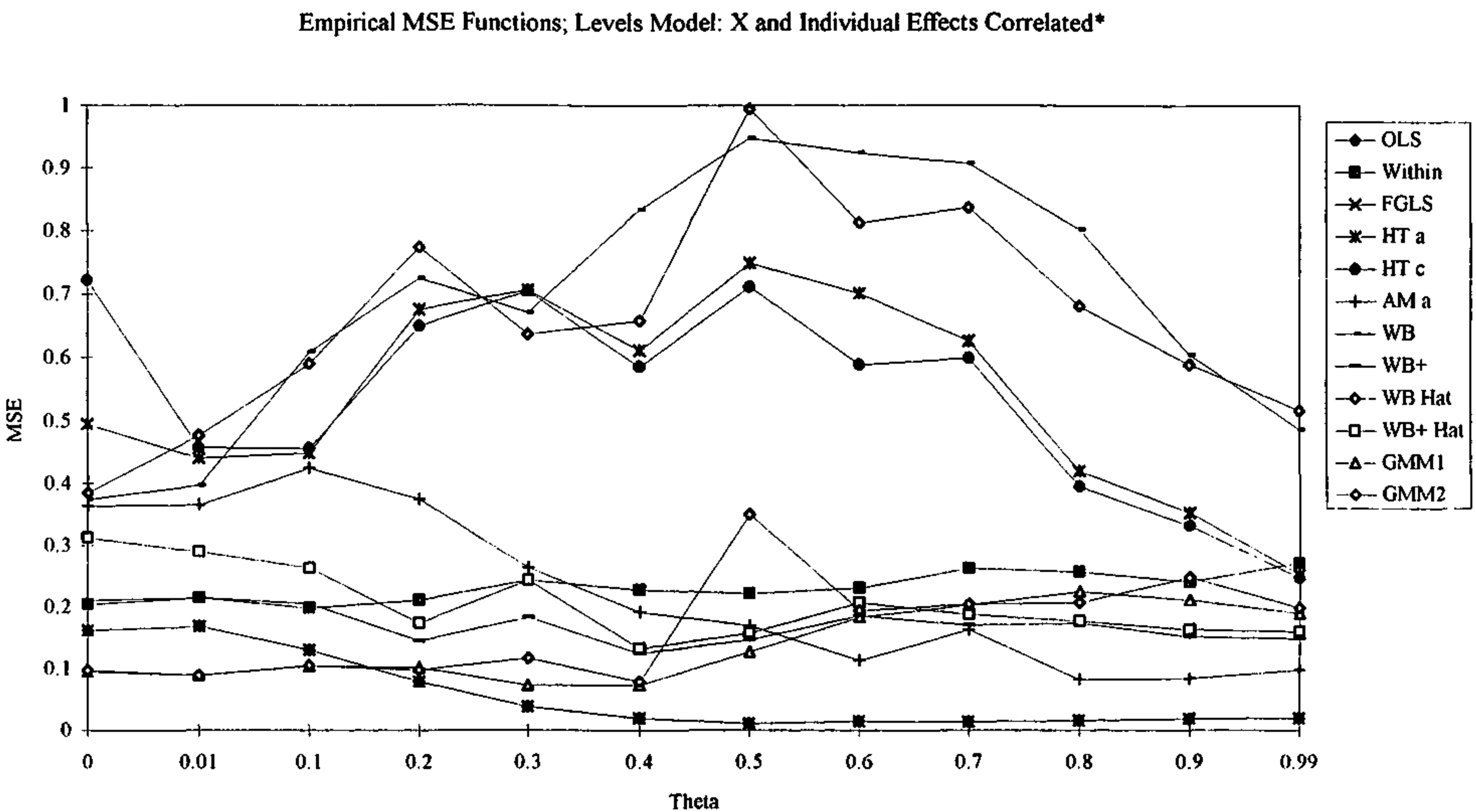
**AB*ov , *AB*ov hat , *HT (a and b)* and *BN (a)* not included due to excessive MSE. *AM (b and c)* not included as performance identical to *AM (a)* .

Figure 4: Random Effects Model; N=25, T=4



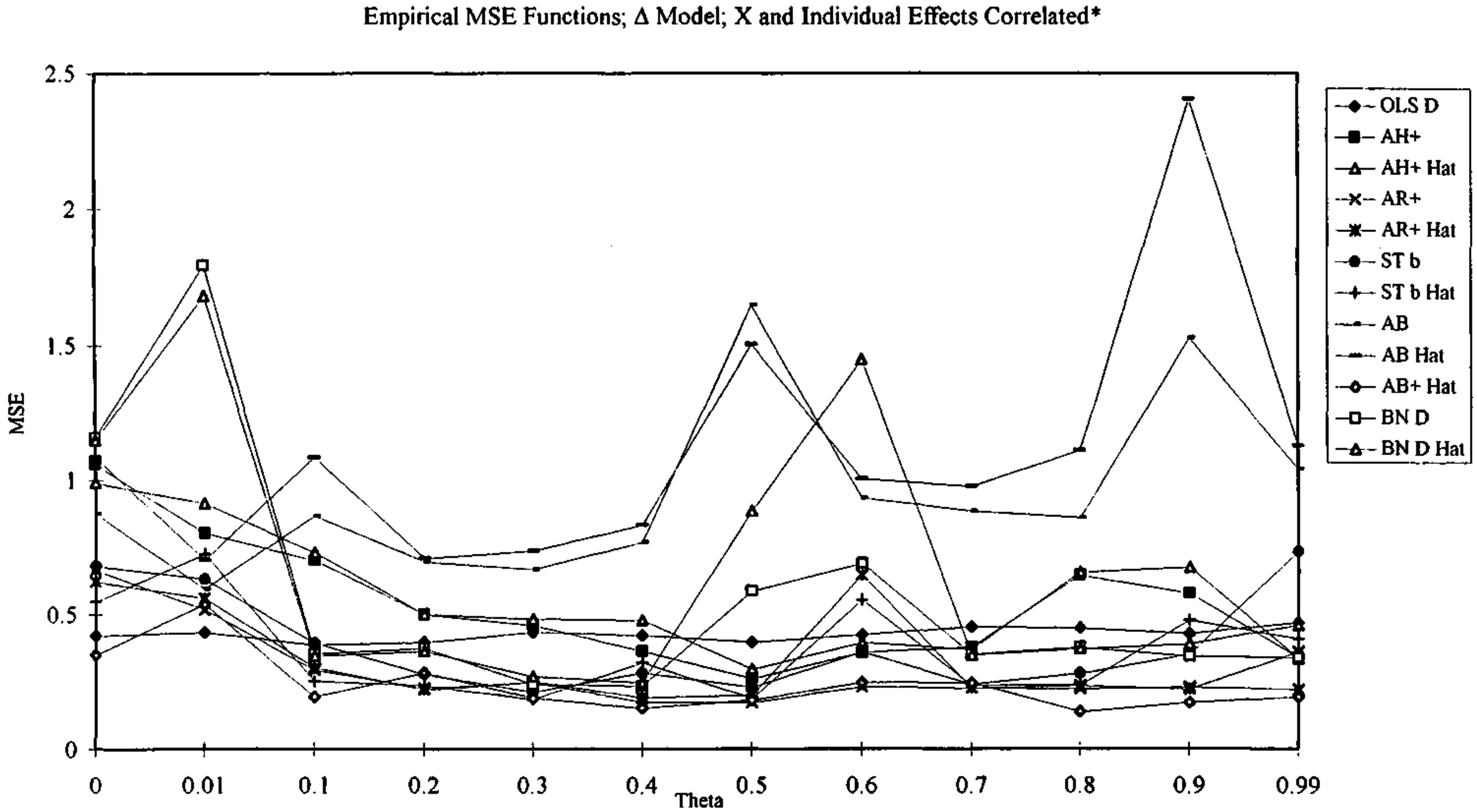
**AR , AH, AH+ hat , AB+ , AB+ hat* and *ST (b) hat* not included due to excessive MSE .

Figure 5: Random Effects Model; N=25, T=4



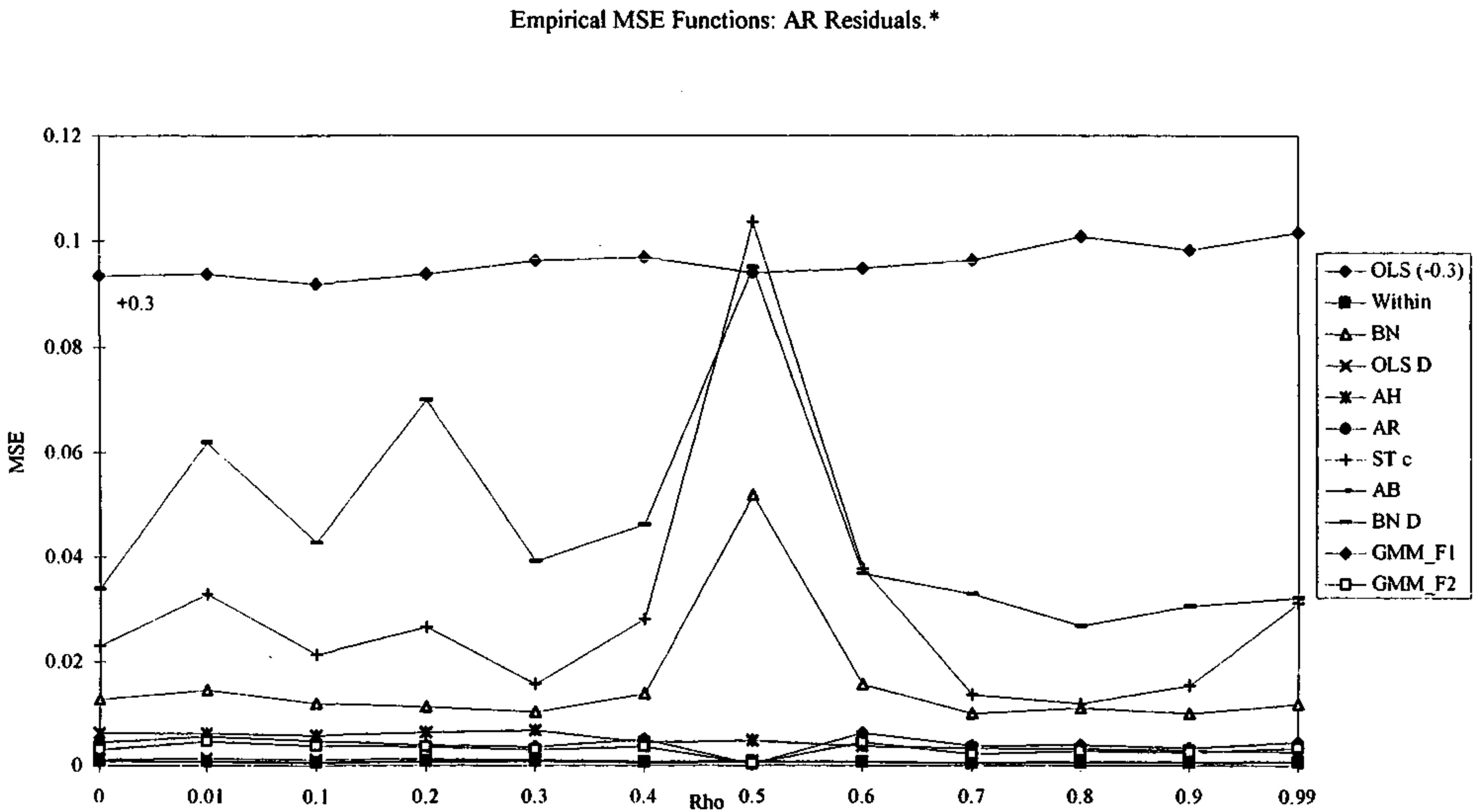
**ABov* , *ABov hat* , *BN (a to c)* and *HT (c)* not included due to excessive MSE. *AM (b to c)* not included as identical performance to *AM (c)* .

Figure 6: Random Effects Model; N=25, T=4



**AR* , *AB+* , *ST (c)* and *AH* not included due to excessive MSE.

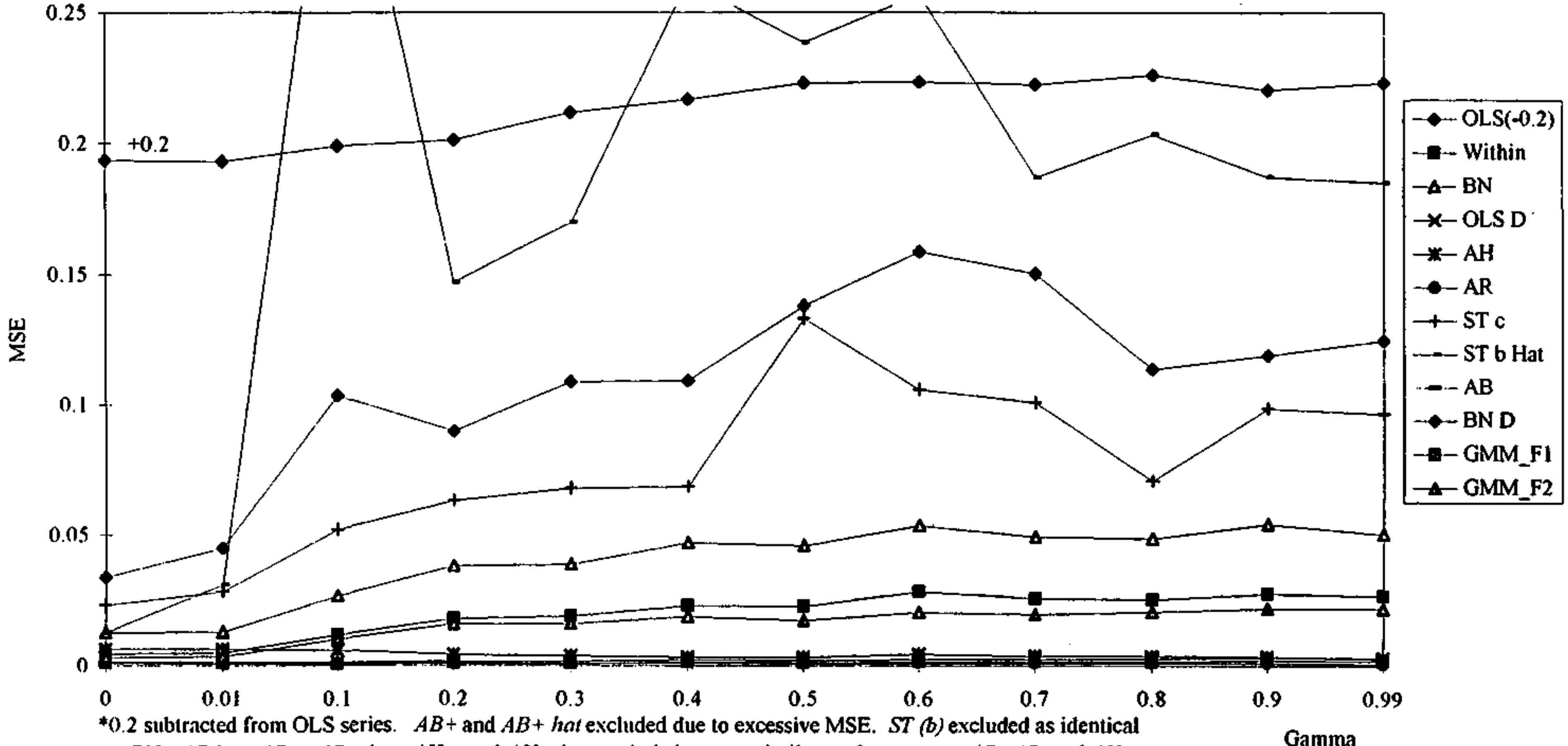
Figure7: Fixed Effects Model; N=25, T=4



*0.3 subtracted from OLS series. *AB+* excluded due to excessive MSE. *AB hat* excluded as identical performance to *AB*. *AH+*, *AH+ hat* and *AR+* and *AR+ hat* excluded as identical performance to *AH* and *AR*, respectively. *ST (b)* excluded as identical to *BN*.

Figure8: Fixed Effects Model; N=25, T=4

Empirical MSE Functions: X and Disturbances Correlated.*



*0.2 subtracted from OLS series. *AB+* and *AB+ hat* excluded due to excessive MSE. *ST (b)* excluded as identical to *BN*. *AB hat*, *AR+*, *AR+ hat*, *AH+* and *AH+ hat* excluded as very similar performance to *AB*, *AR* and *AH*, respectively.