Nonlinear Models and Small Sample Performance of the Generalized Method of Moments

Eva Ventura*

Universitat Pompeu Fabra, (Barcelona)

October 1996

Mailing Address:

Eva Ventura
Universitat Pompeu Fabra
Department of Economics
Ramon Trias Fargas 25-27
08005 Barcelona, SPAIN
Ph: (34)-(3)-5421760
Fax: (34)-(3)-5421746
e-mail:ventura@upf.es

Abstract

In this paper I explore the issue of nonlinearity (both in the data generation process and in the functional form that establishes the relationship between the parameters and the data) regarding the poor performance of the Generalized Method of Moments (GMM) in small samples. To this purpose I build a sequence of models starting with a simple linear model and enlarging it progressively until I approximate a standard (nonlinear) neoclassical growth model. I then use simulation techniques to find the small sample distribution of the GMM estimators in each of the models.

1 Introduction

The Generalized Method of Moments (GMM) estimation technique is intuitively appealing and easy to apply. It is suited for linear or nonlinear models and does not require any distributional assumptions on the disturbances of the model. It only requires orthogonality between the forecast errors and the instruments, and this arises naturally from optimization in Rational Expectations models.

Inference in GMM is asymptotic. For a sufficiently large sample size, the estimators of the model conveniently scaled behave as a Normal vector. It is possible to test the overall fit of the model or to perform Wald type tests on the parameters, but sample sizes are typically small and the asymptotic values are not well approximated. Many papers have studied the small sample distributions of the GMM estimators of particular models. Often they find that these estimators are biased, their asymptotic variance is not correctly estimated and therefore the confidence intervals are wrong, and the models tend to be under or over rejected. In some cases, the estimation results are "perverse": in many Monte Carlo simulations using more instruments yields estimators with lower variances, but a higher mean square error.

The type of models typically estimated by GMM share some common characteristics. Often they are highly nonlinear and this nonlinearity may arise from two sources: the functional form of the equation to estimate and/or the data generating process. The variables involved may exhibit a high degree of autocorrelation or colinearity, or the residuals of the model might be heteroskedastic. I abstract from other issues, such as measurement error in the data, autocorrelation of the

1

residuals which gives rise to alternative ways of estimating the GMM weighting matrix, or several other technical considerations. In this paper I ask myself the following question: is it possible to know to what extent some of these commonly shared characteristics are responsible for the poor performance of the GMM estimation method? If the answer were yes, maybe one could learn more about small sample properties of the GMM estimator from simpler models.

I start by simulating a standard neoclassical growth model. For this model both the data generating process and the derived Euler equations are highly nonlinear. I then extract a collection of many small samples, and derive the small sample distributions of the GMM estimators of the parameters of the model. As expected, the estimators are biased, and they concentrate more and more on the biased values as one keeps adding instruments. I use the data generated by this model to calibrate several of the parameters of the data generating processes in the simulation of a sequence of models. This sequence starts with a simple classical linear regression model and ends with a specification that resembles the Euler equation of the neoclassical growth model. In each step a new characteristic is added or subtracted to the model, keeping constant some of the calibrated data parameters. By simulation, I derive the small sample distribution of the parameters of each model and then I try to establish a relationship between the characteristics of each model and the performance of the GMM estimator. If, for a given model, it is possible to discover the problem that causes the wrong estimates, one could try to find a way to circumvent it or to take it into account to improve estimation.

After examining this sequence of models I have learned that although nonlinea-

rity seems to be mostly responsible for the poor small sample behavior of GMM, it is not easy to disentangle its effect from the effect of the rest of the characteristics such as the degree of autocorrelation of the variables for instance. All of these characteristics interact to yield a particular estimation result, and the way they do it is very model dependent. I could not find a clear pattern and therefore I can not extract some sort of a general set of rules or tests to take into consideration when estimating models with the GMM technique. Instead I have learned that each model has to be studied separately before attempting its estimation and simulation seems to be a very powerful tool to understand the relationship between the model's characteristics and the small sample performance of GMM estimation. One can study the characteristics of the data, calibrate the model he thinks to be correct, and investigate the behavior of the GMM objective function when samples are small or the possibly systematic and model specific relationship between the true parameters values and the bias in their estimation.

The outline of the paper is as follows. First, I review some of the relevant literature in section 2. In section 3, I present an example of GMM estimation applied to a simple growth model. I specify the model, comment on the simulation technique I apply, and state the two stage GMM I use in estimation. Section 4 explores the sequence of models I just mentioned and section 5 concludes the paper.

2   Some common findings in the related literature

There exists some literature that explores the exact small sample properties of instrumental variables estimation in the context of linear models. Nelson and Startz (90) studied them in a linear model in which the instrument was a noisy measure of the explanatory variable and the later was correlated with the error. They found that the central tendency of the instrumental variable estimator was biased away from the true value in the direction of the probability limit of the ordinary least squares estimator. Furthermore, the standard errors associated with these estimates were small relative to the bias. Also, when the instruments were poor (not very correlated to the explanatory variable) the asymptotic distribution was a poor approximation to the true distribution.

Many other papers have used Monte Carlo type strategies to evaluate the small sample properties of IV or GMM estimators. In one of the first articles that looked at the small sample properties of GMM, Tauchen (86) examined the properties of the estimators of utility function parameters. He simulated stochastic nonlinear exchange economies and used them to generate many small samples on which to test the GMM properties. He found a variance/bias trade-off regarding the number of lags used to form instruments (short lags yielded nearly asymptotically optimal estimates of the risk aversion parameter, but long lags tended to produce estimates that concentrated around biased values). That bias depended on the covariance structure of dividends and consumption. The test of the overidentifying restrictions performed well in small samples. If anything, the test was biased towards acceptance of the null hypothesis.

This type of results has been confirmed by other authors such as Chang and Judd (92) in a discrete time stochastic exogenous growth model, Kocherlakota (90) in an asset pricing model similar to the one studied by Tauchen but including the risk free asset and other parameters, or Rogerson and Rupert (93) in a simple real business cycle model with measurement error in hours and wages.

Some papers concentrate on the idea that the GMM weighting matrix is badly estimated and that improving that estimation will result in a better performance of the method. The following are just a few examples. Hansen, Heaton and Yaron (95), Christiano and den Haan (95), Altonji and Segal (94), or Burnside and Eichenbaum (94) explore several alternatives regarding the estimation of the weighting matrix. Although these papers are able to find some improvements, they are in general small and the proposed solutions appear to be model specific or require a minimum data size.

Andersen and Sørensen (95) studied the so-called lognormal stochastic autoregressive volatility model and found a fundamental trade-off between the number of moments, or information, included in estimation and the quality, or precision, of the objective function used for estimation. It is generally not optimal to include a large number of moments in the estimation procedure if the sample size is limited and is virtually never advisable to rely on the alternative extreme of a just-identified model.

Another line of research tries to find a relationship between the quality of the IV used in estimation and the performance of the IV or GMM procedures.

Fuhrer, Moore and Schuh (93) and West and Wilcox (95) studied and compared

ML and GMM estimation in a linear-quadratic inventory model. They concluded that it is the small sample bias that is the cause of so much difference in estimation of the parameters of this model in the relevant literature. FIML gave a larger asymptotic efficiency when the data was highly serially correlated. Also, more efficiency was obtained when the number of instruments in GMM or IV was large, but both methods yielded estimates with the wrong sign and wrong Wald type tests.

Staiger and Stock (93) performed a similar exercise in the context of two stage least squares estimation. They found that even in large samples 2SLS can be badly biased and the conventionally constructed confidence intervals will fail to have the desired coverage rates. Then they compared this estimator with the LIML one. The latter was in many cases approximately median unbiased and therefore more reliable.

Finally, Pagan and Yung (93) advised to perform several previous tests to gain intuition on the small sample performance of the instrumental variables estimators. In particular they suggest the calculation of a pseudo-concentration estimator. This would consist on a regression of the derivatives of the residuals with respect to the parameters of the model, on the instruments one proposes to use. When they apply this test to the Mao's model, they find low values of the R^2 which they interpret as a signal to bad performance of the GMM estimator in small samples. However, they also warn the reader that the results of the tests they propose are not free from contamination from other factors and they are, in general, model dependent.

These are just some examples of work in the field, the list is by no means ex-

haustive. Although summarizing all the related literature in a few lines is very difficult, one can at least select a few common findings. First, it seems that a large number of instruments (lags) improves efficiency of the GMM estimator. But shorter lags (which imply fewer instruments) seem to yield less biased estimates. Second, the J-square statistic behaves somehow unpredictably. Whether the model is overrejected or under rejected depends on the particular model I am studying. Third, the estimated asymptotic variance of the estimators is usually far from the true one and confidence intervals are consequently wrong. Some authors believe that the poor performance of the J-square or Wald tests is due to the bad estimation of the parameters in small samples, but despite the efforts to improve the performance of the estimation method (mainly related to the improvement in the estimation of the weighting matrix), the problem remains unsolved. Finally, the size of the bias or the percentage of rejection of the model (based either on the J-square or Wald tests) appears to be model specific and it could be related to the covariance structure of the data. In this paper, I explore the last issue.

3 An example of GMM estimation: the neo-classical simple growth model

In this section I state a simple stochastic neo-classical growth model (the same used by Rogerson and Rupert ((93)), simulate it for a given choice of parameters, produce 500 samples of 115 observation each, and use them to find the small sample distributions of the parameters estimated by GMM, their estimated asymptotic variance and the J- square statistic.

3.1 The Model

The model is described by

$$\max \mathrm{E}_0 \qquad \sum_{t=0}^{\infty} \delta^t \left(\frac{c_t^{1-\gamma}}{1-\gamma} - A l_t^{\gamma^L} \right) \qquad \gamma > 0 \qquad \gamma^L > 0$$

$$s.t. \quad c_t + k_t = \theta_t k_{t-1}^{\alpha} l_t^{\alpha^L} + \mu k_{t-1}$$

$$\log \theta_t = \rho \log \theta_{t-1} + \epsilon_t \qquad \epsilon_t \sim \mathrm{N}(0, \sigma^2)$$

$$0 \leq l_t \leq 1$$

where

c_t	is consumption at time t
l_t	is hours of work at time t
k_t	is capital stock at the end of period t
θ_t	is a technology shock that follows $\log(\theta_t) = \rho \log(\theta_{t-1}) + \epsilon_t$ with $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$
$1 - \mu$	is the depreciation rate and
$\alpha + \alpha^L$	equal one

The first order conditions for the problem yield:

$$c_t + k_t = \theta_t k_{t-1}^{\alpha} l_t^{\alpha^L} + \mu k_{t-1} \qquad (1)$$

$$c_t^{-\gamma} = \delta \mathrm{E}_t c_{t+1}^{-\gamma} (\theta_{t+1} l_{t+1}^{\alpha^L} \alpha k_t^{\alpha-1} + \mu) \qquad (2)$$

$$l_t = \left(\frac{c_t^{-\gamma} \theta_t \alpha^L k_{t-1}^\alpha}{A\gamma^L} \right)^{\frac{1}{(\gamma^L - \alpha^L)}} \tag{3}$$

The decentralized version of the model implies that the representative agent solves the following problem

$$\max \mathrm{E}_0 \qquad \sum_{t=0}^{\infty} \delta^t \left(\frac{c_t^{1-\gamma}}{1-\gamma} - Al_t^{\gamma^L} \right) \qquad \gamma > 0 \qquad \gamma^L > 0$$

$$s.t. \qquad c_t + B_t \le (1 + r_{t-1})B_{t-1} + w_t l_t$$

$$c_t \ge 0 \quad 0 \le l_t \le 1$$

where r_t is the interest rate, w_t is the wage rate and B_t represents assets holdings at the end of the period t.

The $\{w_t\}$ series is found by making the wage equal to the marginal product of labor. In our case,

$$w_t = \theta_t k_{t-1}^\alpha \alpha^L l_t^{\alpha^L - 1} \tag{4}$$

The natural choice of the interest rate from the model is the risk-free interest rate, which for our specification is given by

$$1 + r_t = (1/\delta) \frac{c_t^{-\gamma}}{\mathrm{E}_t(c_{t+1}^{-\gamma})} \tag{5}$$

The first order conditions for this model yield

9

$$E_t \left\{ \delta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + r_t) \right\} = 1 \qquad (6)$$

or, equivalently:

$$E_t \left\{ \delta \left(\frac{l_{t+1}}{l_t} \right)^{\gamma^L - 1} \frac{w_t}{w_{t+1}} (1 + r_t) \right\} = 1 \qquad (7)$$

3.2 Simulation

The model can be solved numerically. Our particular solution is based on the Parameterized Expectations Approach, formally presented in Marcet and Marshall (94). A sketchy description is offered in the Appendix. The method substitutes the conditional expectation in equation (2) by a suitable function ψ of the state variables k_{t-1} and θ_t and iterates on the parameters of the ϕ function to minimize the distance between the series obtained in successive iterations. Consumption then is given by $c_t^{-\gamma} = \delta \psi(\hat{\beta}, k_{t-1}, \theta_t)$, I substitute it in equation (3) to find l_t, and finally I obtain k_t from equation (1) and the wage from (4). To find the risk-free interest rate, I will have to approximate $E_t c_{t+1}^{-\gamma}$ in the same fashion as before. Then $E_t c_{t+1}^{-\gamma}$ is substituted by $\phi(\hat{\xi}, \theta_t, k_{t-1})$.

The results are dependent on k_0, ρ, var(θ_t), δ, γ, A, γ^L, α, α^L, μ, and the particular draw for $\{\theta_t\}$.

To try to reproduce some of the previous literature results, we choose

$$
\begin{array}{llll}
\delta = & 0.99 & \gamma^L = & 1 \\[4pt]
\mu = & 0.975 & \rho = & 0.95 \\[4pt]
\gamma = & 1 & \sigma^2 = & 0.01034 \\[4pt]
\alpha = & 0.36 & k_0 = & 12.6 \\[4pt]
\alpha^L = & 0.64 & &
\end{array}
$$

A is chosen to be such that given the above parameters, the value of the steady state equilibrium for labor is approximately equal to $\frac{1}{3}$. I assume that the economy is already in the steady state in the initial period and consequently k_0 takes the steady state value for capital.

The stochastic steady state results are shown in table 1. The numbers under the columns labeled c, k, l, θ, w and r are the long-run means and standard deviations of the generated series.

3.3   GMM estimation

Equation (7) implies that

$$
\mathrm{E}\left(u_{t+1}(\delta\,,\,\gamma^L) \otimes z_t\right) = 0
$$

where

$$
u_{t+1}(\delta\,,\,\gamma^L) = \delta\left(\frac{l_{t+1}}{l_t}\right)^{\gamma^L-1}\frac{w_t}{w_{t+1}}(1+r_t) - 1
$$

and z_t is a vector of valid instruments, that is of variables belonging to the

information set at time t.

Our GMM estimators of δ and γ^L minimize

$$\left(\frac{\sum_{t=1}^{T} u_{t+1}(\delta \, , \, \gamma^L) \otimes z_t}{T} \right)' (S_{wT})^{-1} \left(\frac{\sum_{t=1}^{T} u_{t+1}(\delta \, , \, \gamma^L) \otimes z_t}{T} \right)$$

where the inverse of the weighting matrix is defined as:

$$S_{wT} = \frac{\sum_{t=1}^{T} \left(\overline{\delta}_T \left(\frac{l_{t+1}}{l_t} \right)^{\overline{\gamma^L}_T - 1} \frac{w_t}{w_{t+1}} (1 + r_t) - 1 \right) \otimes z_t z_t \prime}{T} \qquad (8)$$

and $\overline{\delta}_T$ and $\overline{\gamma^L}_T$ are first stage estimates of the parameters.

The simplest method calls for using the identity matrix as a first stage estimation weighting matrix. The first stage estimates are used to evaluate (8) and second stage estimators are obtained. It is possible to improve the estimation by repeating the iteration process several times, until convergence of the estimators is achieved (see Hansen, Heaton and Yaron (95)). However, for the particular model I am studying the estimators are still biased and the fit of the model is not generally improved[1].

Asymptotically,

$$\sqrt{T} \left(\begin{pmatrix} \delta_T \\ \gamma^L_T \end{pmatrix} - \begin{pmatrix} \delta \\ \gamma^L \end{pmatrix} \right) \xrightarrow{\mathcal{D}} N \left(0, \left(B'(S_w)^{-1} B \right)^{-1} \right)$$

where B is estimated by

$$B_T = \frac{\sum_{t=1}^{T} \left(\left(\frac{l_{t+1}}{l_t}\right)^{\gamma_T^L - 1} \frac{w_t}{w_{t+1}}(1 + r_t) \ , \ \delta_T \left(\frac{l_{t+1}}{l_t}\right)^{\gamma_T^L - 1} \frac{w_t}{w_{t+1}}(1 + r_t)\ln\left(\frac{l_{t+1}}{l_t}\right) \right) \otimes z_t}{T}$$

$$(9)$$

Also

$$J_T = T \left(\frac{\sum_{t=1}^{T} u_{t+1}(\delta_T \ , \ \gamma_T^L) \otimes z_t}{T} \right)' (S_{w_T})^{-1} \left(\frac{\sum_{t=1}^{T} u_{t+1}(\delta_T \ , \ \gamma_T^L) \otimes z_t}{T} \right)$$

converges in distribution to a χ^2 variable with as many degrees of freedom as the number of overidentifying restrictions of the model.

3.4   Small Sample Properties

In order to characterize the small sample properties of the GMM estimator, we generate 500 series of 115 observations each[2].

For each one of these samples, I estimate the δ and γ^L parameters with several sets of instruments, which I choose among the variables that will be typically available when estimating this type of models. These sets of instruments are shown in table 2.

The results of estimating the model with each set of instruments are shown in table 3.

3.5   Comments

Table 3 can be summarized in a few points. First, the bias of the parameters' estimators is lower with the first set of instruments, the one I call Lag 0. Second, the

small sample standard error of the parameters' estimators is systematically larger than the asymptotic one, and both decrease as the number of instruments (lags) increases. Third, the square root of the mean square error of the parameters' estimators increases with the number of instruments (lags) being used in estimation. Fourth, the model's overidentifying restrictions are always rejected at the 10% confidence level[3]. The percentage of rejects decreases with the number of instruments being used. Fifth, as noted before by Rogerson and Rupert (93) the estimator of δ displays very little bias.

These results are somehow "perverse". If the econometrician was to choose among several sets of instruments, since he does not know the true value of the parameters and all he sees is the asymptotic standard error estimate and the J-statistic, he would undoubtedly prefer to include as many instruments as possible. Our simulations show that unfortunately this would result in quite biased estimators.

4   The role of nonlinearity

The preceding section simply documents a well known fact: usually (but not always) GMM does not perform too well with small samples. A recurrent finding in the related literature is that the performance of GMM estimation and the possibility of success in improving it, is model dependent.

Something is known about the behavior of instrumental variables estimation in the context of linear models (see Amemiya (66) or Nelson and Startz (90)). In particular, a lot of attention has been devoted to studying the effects of using

poor instruments in IV estimation. Instruments that are lowly correlated with the explanatory variables tend to produce biased estimators. But this does not seem to be the problem in the model I just estimated. To the contrary, the instruments are highly autocorrelated and the correlation among some of them is also large. Under that circumstances, simple linear models tend to perform well.

That observation suggests that one could learn on the issue of nonlinearity (either in the data generation process or in the relationship between data and parameters) by considering a sequence of models, starting with a linear one and ending with an approximation to the growth model I just estimated.

4.1   The Models

All the models include some degree of autocorrelation in the variables, which is common in Rational Expectations models. Such autocorrelation is absolutely necessary if one is going to use lagged values of the variables as instruments.

Sometimes the autocorrelation structure is explicit in the equation one wants to estimate, such as in models 4 to 7. In the other three cases the data generation process of the autocorrelated variable is independent of the structure of the model. I pay special attention to the role played by nonlinearity, both in the functional form of the equation being estimated and in the data generating process, and how it interacts with other characteristics of the models. Model 1 is linear in the variables and in the parameters, while models 2 to 6 are nonlinear in the parameters, and model 7 is nonlinear both in variables and parameters. I distinguish between models in which the parameters enter in a nonlinear form just in the right hand side of

15

the equation to be estimated (models 2 and 4), and models where that type of nonlinearity is also present in the left hand side of the equation. The explanatory variables in models 1 to 3 are correlated with the residuals of the model, as opposite to the rest of the models. Models 6 and 7 include heteroskedastic residuals.

The sequence of models that I consider is the following:

Model	Functional form	Data Generating Process		
1	$y_t = \beta x_t + u_t$	$x_t = \omega_t + \lambda u_t$	$u_t \sim N(0, \sigma_u^2)$	
		$\omega_t = \rho\omega_{t-1} + u_t^\omega$	$u_t^w \sim N(0, \sigma_{u^w}^2)$	
2	$y_t = x_t^\beta + u_t$	$u_t = \exp(\nu_t) - \exp\left(\frac{\sigma_\nu^2}{2}\right)$	$\nu_t \sim N(0, \sigma_\nu^2)$	
		$\omega_t = \rho\omega_{t-1} + u_t^\omega$	$u_t^w \sim N(0, \sigma_{u^w}^2)$	
		$x_t = \exp(\omega_t + \lambda u_t) +$	$\left(\exp\left(\frac{\sigma_\nu^2}{2}\right)\right)^{\frac{1}{\beta}}$	
3	$y_t^{\beta+\gamma} = x_t^\beta + u_t$	$u_t = \exp(\nu_t) - \exp\left(\frac{\sigma_\nu^2}{2}\right)$	$\nu_t \sim N(0, \sigma_\nu^2)$	
		$\omega_t = \rho\omega_{t-1} + u_t^\omega$	$u_t^w \sim N(0, \sigma_{u^w}^2)$	
		$x_t = \exp(\omega_t + \lambda u_t) +$	$\left(\exp\left(\frac{\sigma_\nu^2}{2}\right)\right)^{\frac{1}{\beta}}$	
4	$x_t = x_{t-1}^\beta + u_t$	$u_t \sim N(0, \sigma_u^2)$		
5	$x_t^{\beta+\gamma} = x_{t-1}^\beta + u_t$	$u_t \sim N(0, \sigma_u^2)$		
6	$y_t^{\beta+\gamma} = x_{t-1}^\beta + \tilde{u}_t$	$x_t = x_{t-1}^\nu + u_t$	$y_t = \frac{x_t}{x_{t-1}^{\nu-\beta}}$	
		$u_t \sim N(0, \sigma_u^2)$	$\tilde{u}_t = \frac{u_t}{x_{t-1}^{\nu-\beta}}$	
7	$\delta\left(\frac{x_t}{x_{t-1}}\right)^\beta \psi_t - 1 = \tilde{u}_t$	$\psi_t = \left(\frac{x_t^{1-\beta} x_{t-1}^{-\nu}}{\delta x_{t-1}^{-\beta}}\right)$	$\tilde{u}_t = \frac{u_t}{x_{t-1}^\nu}$	
		$x_t = x_{t-1}^\nu + u_t$	$u_t \sim N(0, \sigma_u^2)$	

There is only one parameter to estimate through all the models, β. Its true

value is set to 0.9 through all the experiments and I also try to keep fixed some of the rest of the parameters that govern the statistical properties of the explanatory variables. In particular, if possible I will try to match the autocorrelation observed in consumption in the model of section 3 (0.994), its standard deviation (0.0414), and the standard deviation of the forecast errors (0.0047)[4]. The γ parameter is set to 0.1, thus making models 2 and 3 equivalent as well as models 4 and 5.

I will estimate each model with several sets of instruments. The instruments will be divided in two classes: instruments which are noisy versions of the relevant ones, and instruments which are lagged versions of the relevant ones.

For models 1 to 3 I define the variables:

$$z_{1t} = w_t + v_{1t}$$

$$z_{2t} = w_t + v_{2t}$$

$$z_{3t} = w_t + v_{3t}$$

$$z_{4t} = x_{t-1}$$

where v_{1t}, v_{2t}, and v_{3t} are i.i.d. noises with mean 0 and standard deviation equal to 0.15 times the standard deviation of w_t.

I will use the following instruments sets:

IV Set	Variables
1	z_{1t}
2	z_{1t}, z_{2t}, z_{3t}
3	z_{1t}, z_{2t}, z_{3t}, $z_{1,t-1}$, $z_{2,t-1}$, $z_{3,t-1}$
4	z_{1t}, z_{2t}, z_{3t}, $z_{1,t-1}$, $z_{2,t-1}$, $z_{3,t-1}$, $z_{1,t-2}$, $z_{2,t-2}$, $z_{3,t-2}$
5	$z_{4,t}$
6	$z_{4,t}$, $z_{4,t-1}$
7	$z_{4,t}$, $z_{4,t-1}$, $z_{4,t-2}$, $z_{4,t-3}$
8	All of the above

For models 4 to 6, I redefine the variables:

$$z_{1t} = x_t + v_{1t}$$

$$z_{2t} = x_t + v_{2t}$$

$$z_{3t} = x_t + v_{3t}$$

$$z_{4t} = x_{t-1}$$

where now v_{1t}, v_{2t}, and v_{3t} are i.i.d. noises with mean 0 and standard deviation equal to 0.15 times the standard deviation of x_t.

The instrument sets used with models 4 to 6 are the same as the ones used for models 1 to 3, taking into account the redefinition of z_{1t}, z_{2t}, z_{3t} and z_{1t}.

Finally, the instruments sets for model 7 are:

IV Set	Variables
1	x_{t-1}
2	x_{t-1}, x_{t-2}
3	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$
4	ψ_{t-1}
5	ψ_{t-1}, ψ_{t-2}
6	$\psi_{t-1}, \psi_{t-2}, \psi_{t-3}, \psi_{t-4}$
7	x_{t-1}, ψ_{t-1}
8	$x_{t-1}, x_{t-2}, \psi_{t-1}, \psi_{t-2}$
9	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, \psi_{t-1}, \psi_{t-2}, \psi_{t-3}, \psi_{t-4}$

4.2 Calibration and Data Statistics of the Models

4.2.1 Model 1

Here I make $\rho = 0.994$ to match the autocorrelation of consumption in the growth model of section 3, $\sigma_{u^w} = 0.0046$ to match its standard deviation, and $\sigma_u = 0.0047$ to match the standard deviation of the forecast error. The presence of λu_t in the definition of x_t justifies the use of instruments to estimate the β parameter. The reference value for λ is 0.9, the one used by Nelson and Startz (90) in a very similar model. I also use alternative values for this parameter, to better evaluate in which way the estimation results are affected by the degree of correlation between the instruments and the explanatory variable.

With these choices of parameters, the data statistics are as shown in table 4.

The numbers under the columns labeled T=10,000 are estimates of the popu-

lation values of the reported statistics. Here σ_u and σ_x represent the standard deviation of u_t and x_t respectively. The autocorrelation coefficients for u_t and x_t are represented as ρ_u and ρ_x respectively. The symbol $\rho_{u,x}$ refers to the correlation between u_t and x_t. The numbers under the columns labeled T=115(500) are small sample statistics. They report the means over 500 realization each of one is of length 115. As can be seen from the table, the standard deviation of x is systematically larger in the small samples while its autocorrelation is slightly smaller. The correlation between x and u is also larger in the small samples.

4.2.2 Model 2

Again, $\rho = 0.994$ and I make $\sigma_\nu = 0.0047$ and $\sigma_{u^w} = 0.00456$. That way I am able to match the autocorrelation and standard deviation of consumption. I also match the standard deviation of the forecast error.

With this choice of parameters, the data statistics are as shown in table 5.

It should be noted that the functional form of the model is not the only change performed. In order to guarantee stability of the model, the data generating process is also different. In particular, the x_t variable is lognormal.

4.2.3 Model 3

Here I make $\gamma = 0.1$, therefore $\beta + \gamma = 1$. The rest of the parameters take the same values as in model 2. Both models are -for this particular choice of parameters- identical from the point of view of data generation.

20

4.2.4 Model 4

For this particular model I can not match both the autocorrelation and the standard deviation of consumption. If I choose $\sigma_u = 0.01795$ I match the standard deviation of consumption, but its (linear) autocorrelation is low (0.9) and the standard deviation of the residuals is too big. Making $\sigma_u = 0.0047$ as before implies that the standard deviation of x_t is only 0.01083, and its autocorrelation is still 0.9. I will chose then $\sigma_u = 0.01795$. The small sample statistics associated to this choice are 0.0389 for the standard deviation of x_t and 0.8866 for its (linear) autocorrelation.

4.2.5 Model 5

I make $\gamma = 0.1$ and the rest of the parameters take the same values as in the previous model. Both are identical from the point of view of data generation.

4.2.6 Model 6

I here separate the data generation mechanism from the model functional equation. I make $\rho = 0.994$ and $\sigma_u = 0.00458$ so I can match the autocorrelation and standard deviation of consumption. The small sample statistics are 0.0205 for the standard deviation of x_t and 0.9727 for its (linear) autocorrelation.

4.2.7 Model 7

The model is a transformation of the data generation equation for x_t, just like model 6. Again the ν parameter is equal to 0.994. The δ parameter is equal to 0.99 and $\sigma_u = 0.00458$. The small sample statistics are 0.02054 for the standard deviation of x_t and 0.9727 for its (linear) autocorrelation coefficient.

I have also estimated the same model with a lower autocorrelation of x_t. For $\nu = 0.194$ the small sample statistics are 0.00464 for the standard deviation of x_t and 0.8193 for its (linear) autocorrelation coefficient.

4.3 Results

Tables 6 to 8 gather the estimation results for the sequence of models. Sometimes I try alternative choices of some parameters , to learn more about the particular structure of the model and its interaction with GMM.

The tables report the small sample estimated bias , the standard deviation and the square root of the mean square error of the estimator. Also, I report the percentage of rejects of the $J-$statistic test at the 10% confidence level. Finally I include the estimated asymptotic standard deviation of $\sqrt{T}\beta_T$ divided by $\sqrt{T}$ when T=115 ($(115)^{-1/2}\sigma_{\beta_T}$ in the tables), and an approximation to the true asymptotic standard deviation of $\sqrt{T}\beta_T$ as T goes to ∞, divided by 115 ($(115)^{-1/2}\sigma_\beta$ in the tables)[5]. This approximation is calculated by computing the mean of $(B_T\prime S_{w_T}^{-1}B_T)^{-1}$ with T=10,000 over 150 realizations and then dividing it by 115 and taking the square root.

The bias in model 1 is small, even when the quality of the instruments is poor ($\lambda = 10.0$). The small sample standard deviation of the parameter is a little larger, usually around 0.02. However, when $\lambda = 10.0$ and only lagged values of x_t are used as instruments, that standard deviation is considerably larger, with a maximum at 0.39. The finding is explained by the fact that these instruments are very noisy versions of the explanatory variable and therefore only weakly correlated with it.

I also observe a trade-off between bias and variance: when λ is 0 or 0.9 the more instruments I use, the lower the bias and the higher the small sample standard deviation. However, when $\lambda = 10.0$ the result is reversed.

In this model the small sample variance is larger than the estimated asymptotic variance[6], independently of the value of λ. Also, the latter is larger than the approximated true variance of the GMM estimate.

As for the value of the J-statistic, I observe that when λ is 0 or 0.9 the percentage of times the model is rejected is very close to the expected 10%. The percentage diminishes as I include more and more instruments. For $\lambda= 10$, the model is underrejected with instruments sets 2, 3 or 4, sometimes rather badly. But when I use sets 6 and 7 the model is overrejected very often and the situation becomes worst as I increase the number of instruments used.

Models 2 and 3 work remarkably well. The bias is very small. The small sample, estimated asymptotic and approximated true asymptotic standard deviations are very similar and very small, and the percentage of rejects of the J-statistic test is very close to the 10% level. Surprisingly enough, the bias and standard deviations of model 3 are approximately equal to 10 times the ones reported for model 2. Both models are identical from the point of view of data generation, but not from the point of view of estimation. Two are the main differences with model 1. First, there is a nonlinear function of the β parameter on the left hand side of the equation to be estimated. Second, the x_t variable is now lognormal, while in model 1 it was normal.

Models 4 and 5 are purely autoregressive, with the β parameter entering in

a nonlinear fashion. Both models are identical from the point of view of data generation. Table 7 shows the results of the GMM estimation for these two models. I emphasize a few points:

- The bias is large now, specially for model 5.

- The small sample standard deviation is also very large in both models.

- In model 4 the small sample deviation and the estimated asymptotic standard deviation are very close, with the exception made for instrument sets 1 and 5. In model 5 the small sample standard deviation is much smaller than the estimated asymptotic one.

- The trade-off observed before between bias and variance is not observed everywhere. In model 4 the bias and the estimated standard errors diminish as I increase the number of instruments. The same is true for model 5 and instruments sets 1 to 4, but the trade-off reappears with instrument sets 5 to 8.

- In general, the models are estimated very badly when only one instrument is used (sets 1 and 5).

Model 6 shows some important bias and large small sample variances. These are usually larger than the estimated asymptotic standard deviation (exception made of instrument sets 1 and 5), but smaller than the approximated true one. It exhibits the trade-off between bias and variance as the number of instruments increases. Again, the results are meaningless when only one instrument is used in estimation.

The three models are always underrejected for what respects to the J-statistic test. I could never reject the hypothesis of $\beta = 0.9$, but the estimated asymptotic variance is so large that the estimation results offer very little information on the true value of that parameter.

Model 7 extremes the nonlinearity aspect of this type of setups. When I match the sample data values of consumption ($\nu = 0.994$), the estimates are badly biased and the estimated asymptotic standard deviation is large (larger than the small sample one). But it still underestimates the "true" asymptotic standard deviation. The model is underrejected or overrejected according to the J-statistic test, depending on the particular instrument set considered.

Summarizing, one can observe that linear models do not pose estimation problems unless the instruments are poorly correlated with the explanatory variables. Also, nonlinear models of the type explored work well with small samples. The high autocorrelation in the data does not seem to interfere with the quality of the small sample estimators.

GMM has problems to estimate the pure autocorrelation model 4. Although models 5 and 4 are identical from the point of view of data generation, the functional form in model 5 makes estimation very difficult. The large estimated and approximated standard errors suggest that the objective function (both in the small and the large sample) are rather flat. Figures 1 and 2 show the logarithms of T times the objective functions of model 1 (for $\lambda = 0.9$) and models 4 to 7 for different values of T. Figure 1 uses only lagged values of the explanatory variables as instruments (instrument set 7). Figure 2 uses only one instrument, the one period

lagged value of the explanatory variable. It is also easy to appreciate that these functions are sometimes degenerate when only one instrument is used.

Separating the data generating process from the functional form of the model, as in model 6, does not solve the problem posed by the autocorrelation of x_t, but makes the objective function better defined. Models 7 and 6 are also identical from point of view of data generation. However, the particular functional form makes the objective function ill behaved.

In the case of model 7 it can also be observed that when ν is set to 0.194, the bias diminishes although is still quite large. Now the estimated asymptotic deviation is considerably larger than the approximated true one. The model's overidentifying restrictions are now underrejected always. Again, the model can't be estimated properly when just one instrument is used. Figure 1 help us understand why is so. The last three graphics show that the high autocorrelation combined with the particular functional form produce ill behaved objective functions (both for small and large samples). Lowering the autocorrelation or transforming the model help improve these objective functions and therefore give better (although still not satisfactory) estimations.

Figure 3 shows the relationship between the true parameter value and the mean of its GMM estimate over 500 data replications. The models studied are models 5, 6 and 7, all of them with a high autocorrelation in x_t. I have worked with instrument sets 5 and 7 for models 5 and 6 , and with instrument sets 3 and 4 for model 7. The graphics on the left in Figure 3 correspond to cases in which the models are just identified. The graphics on the right correspond to the case in

which the instrument set includes four lags of the x_t variable. Casual examination of these pictures reveals that, in some cases, there is a low variance relationship between the true parameter and the small sample mean of its GMM estimator. I can disregard model 5, which I already know is badly behaved. Models 6 and 7 offer the possibility of estimating a particular functional form relating the true parameter and the mean GMM estimate. For model 5 this function is not defined for a true parameter value of 1.0, but is defined elsewhere. In model 7 it seems quite evident that a straight line describes the relationship between parameter and estimator. There is too much variance when I use instrument set 4, for which the model is just identified. But when instrument set 3 is used, I can estimate a linear regression between the GMM mean estimate and the true parameter for model 7. The R^2 is always very high. In this particular case, I could improve estimation by adding a third step to GMM. For example, for instrument set 3, the estimated linear regression is $\beta_{T}\prime = -1.012555 + 1.001963\beta$, with an adjusted R^2 of 0.9999. Here $\beta_{T}\prime$ is the small sample mean of the GMM estimator, obtained by simulation. I could define a third step GMM estimator as $\hat{\beta}_T = \frac{\beta_T + 1.012555}{1.001963}$, where β_T is the standard two steps GMM estimate. Of course, the asymptotic theory associated with this estimator needs to be derived.

5 Concluding Remarks

In this paper I have tried to establish some relationship between several characteristics present in Rational Expectations models, such as nonlinearity, cross and autocorrelation of the data processes or heteroskedasticity in the forecast errors;

and the performance of the generalized method of moments in small samples.

The results of such effort are somehow disappointing since I have learned that although nonlinearity seems to be mostly responsible for the poor small sample behavior of GMM, it is not easy to disentangle its effect from the effect of the rest of the characteristics. All of them interact to yield a particular estimation result, and the way they do it is very model dependent. I could not find a clear pattern and therefore it is not possible for us to extract some sort of a general set of rules to take into consideration when estimating models with the GMM technique. In some cases, such as in models 6 and 7, a transformation of the model makes an important difference in estimation, but it is not obvious at all which transformation one should try. Also, although lowering the autocorrelation improved estimation in model 7[7], I can not just change the characteristics of the data in estimation. Some voices have pointed out that models such as 7 yield wrong estimates when one uses aggregated data because there is a high correlation between x_t/x_{t-1} and ψ_t. Therefore, it seems reasonable to try to estimate these models with panel data, when this is possible. Panel data will show more dispersion and could improve estimation, but is sometimes subject to measurement error. Furthermore, the objective function changes since I then need as many moment equations as time periods in the data so it is not clear what the final outcome would be. Although I have not tried to simulate a panel here, I have tried to break the high correlation between x_t/x_{t-1} and ψ_t in model 7 by adding an i.i.d. shock to x_t/x_{t-1}. I tried with several variances for the shock, but estimation did not improve significantly or it worsened.

The poor small sample performance of GMM is a fact and very little can be

done to avoid it. However, some further exploration of the results and the literature leads us to conclude that it may be possible to correct the GMM estimations afterwards. The bootstrap method has already been applied in the GMM context in a recent paper by Hall and Horowitz (96), but this method can sometimes be very computationally intensive, especially with the autocorrelated data typically present in the models I have explored here. Its success will still depend to some extent on the size of the sample.

The preliminary graphical exploration of Figure 3 seem to suggest that, in the cases in which the relationship between the true parameter and its GMM estimate is smooth, it may be worthwhile to exploit the parametrics of the models to remove the bias and improve the confidence intervals in a cheap and reliable way. This is the subject of further theoretical research that will be attempted in a sequel to this paper.

APPENDIX 1

PEA algorithm

I describe here how to apply PEA to simulate the model in section 3. Given the parameters of the model and a starting value of capital, I find $\left\{\overline{c}_t, \overline{w}_t, \overline{l}_t, \overline{l}_t, \overline{k}_t, \overline{r}_t\right\}$ that satisfy equations (1) to (7) for all t.

- Step 1; substitute the conditional expectation in the right side of (6) and the conditional expectation in (5) by flexible functional forms of the state variables of the model to obtain

$$u'(\overline{c}_{1,t}) = \delta \ \psi_1(\beta; \overline{k}_{t-1}, \theta_t) \qquad (10)$$

and

$$1 + r_t = (1/\delta)\frac{\overline{c}_t^{-\gamma}}{\psi_2(\xi; \overline{k}_{t-1}, \theta_t)} \qquad (11)$$

Here, I choose ψ_1 and ψ_2 as exponentiated polynomials that are insured to take on only positive values; the parameters β and ξ are the parameters in the polynomials. Fix β and ξ.

- Step 2. Obtain a long simulation $\{\overline{c}_t(\beta, \xi), \ \overline{l}_t(\beta, \xi), \ \overline{w}_t(\beta, \xi), \ \overline{r}_t(\beta, \xi), \ \overline{k}_t(\beta, \xi)\}_{t=0}^{T}$, consistent with these parameterized expectations for large T^8. This is done by, in each period, for given state variables, obtaining $\overline{c}_t(\beta, \xi)$ from the parameterized version of (6), $\overline{l}_t(\beta, \xi)$ from (3), $\overline{w}_t(\beta, \xi)$ from (4),

30

$\overline{r}_t(\beta, \xi)$ is obtained from (11); finally, $\overline{k}_t(\beta, \xi)$ is obtained from (1) and I can move to the next period.

- Step 3. Perform a non-linear regression of $u'(\overline{c}_{1,t+1}(\beta, \xi))$ $((\overline{r}_{t+1}(\beta, \xi) + 1)$ (the expression inside the conditional expectation in (6) on the functional form $\psi_1(\cdot; \overline{k}_{t-1}(\beta, \xi), \theta_t)$. and of $\overline{c}_{1,t+1}(\beta, \xi)$ (the expression inside the conditional expectation in (5) on the functional form $\psi_2(\cdot; \overline{k}_{t-1}(\beta, \xi), \theta_t)$. Call the result of these regressions $G(\beta, \xi)$.

- Step 4. Iterate on β and ξ to find $(\beta_f, \xi_f) = G(\beta_f, \xi_f)$.

The approximate solution is given by $\Big\{\overline{c}_{j,t}(\beta_f, \xi_f), \overline{l}_t(\beta_f, \xi_f), \overline{w}_t(\beta_f, \xi_f), \overline{r}_t(\beta_f, \xi_f), \overline{k}_t(\beta_f, \xi_f)\Big\}_{t=0}^{T}$

This is a very simple version of the method. I only need long run stochastic steady state paths for the variables and I assume that the initial capital stock is already in the stochastic steady state support. The PEA method can encompass other situations, in particular it can be used when one must approximate the transition path of the variables to the steady state very accurately. The reader is referred to (94) and (94) for a formal presentation of the model, to (94) for a discussion in accuracy of simulations with the PEA method, or (92), (96), and (93) for applications.

Endnotes

1. Most of the time the value of the objective function depends on the units being used. For instance, if I compare the estimation results from using a particular set of instruments in one hand and their logarithms in the other, the estimated values of the parameters and their asymptotic variances are very similar but the percentage of rejects of the model is much lower when I use the logarithmic instruments. At least for this particular model, that difference disappears when I use an iterative method to evaluate the weighting matrix instead of the simplest one described and adopted here.

2. Rogerson and Rupert (93) also use 115 observations. Kocherlakota (90) uses 90 and Tauchen (86) generates series of length 50 or 75.

3. In fact I have learned that the model is very sensitive to the units of the instruments. When I use levels on the instruments I get over rejection. If I use logarithms I sometimes get under rejection. The units effect can be removed when I use an iterative method to compute de weighting matrix. Then the overall result is over rejection.

4. All these numbers have been calculated with 10,000 simulated data points.

5. Note that $(B_T \prime S_{w_T}^{-1} B_T)^{-1}$ gives us the estimated asymptotic variance of $\sqrt{T}\beta_T$. To find confidence intervals for β_T, I have to divide that number by T to get the estimated asymptotic variance of β_T. When T takes different values, such as in our exercise, the estimates of $\sqrt{T}\beta_T$ are of similar magnitude, but this is not the case when I divide by T. This is why I have chosen to report

$\sqrt{T}\beta_T(115)^{-1}$ in both cases.

6. With an exception for instrument set 5 and $\lambda = 10.0$. In this case, a few realizations give very bad estimations, and the means reported in the table reflect just that.

7. In fact, only the risk-free interest rate depends both on β and ξ, since it is a function of consumption. The rest of the variables do not depend on r_t and therefore they not depend on ξ.

8. I tried the same strategy with models 5 and 6, but the estimation did not improve and even worsened.

Table 1: **Steady State Distribution (T=10,000)**

	c_t	k_t	l_t	r_t	w_t
Mean	0.9202	12.6354	0.3335	1.0100	2.3760
Std. Dev	0.0414	0.7964	0.0115	0.0015	.1068
Correlations					
w_t ,w_{t-1}	0.994				
l_t ,l_{t-1}	0.895				
r_t ,r_{t-1}	0.908				
w_t ,l_t	0.337				
l_t ,r_t	0.866				
r_t ,w_t	-0.178				

Table 2: **Lists of instruments for the neoclassical growth model.**

Instrument Set	Variables
Lag 0	$1, l_t , w_t, r_t$
Lag 1	$1, l_t , w_t ,r_t, l_{t-1} , w_{t-1}, r_{t-1}$
Lag 2	$1, l_t , w_t ,r_t, l_{t-1} , w_{t-1}, r_{t-1}, l_{t-2} , w_{t-2}, r_{t-2}$
Lag 3	$1, l_t , w_t ,r_t, l_{t-1} , w_{t-1}, r_{t-1}, l_{t-2} , w_{t-2}, r_{t-2}, l_{t-3} , w_{t-3}, r_{t-3}$
Lag 4	$1, l_t , w_t ,r_t, l_{t-1} , w_{t-1}, r_{t-1}, l_{t-2} , w_{t-2}, r_{t-2}, l_{t-3} , w_{t-3}, r_{t-3}$ $l_{t-4} , w_{t-4}, r_{t-4}$
Lag 5	$1, l_t , w_t ,r_t, l_{t-1} , w_{t-1}, r_{t-1}, l_{t-2} , w_{t-2}, r_{t-2}, l_{t-3} , w_{t-3}, r_{t-3}$ $l_{t-4} , w_{t-4}, r_{t-4}, l_{t-5} , w_{t-5}, r_{t-5}$

Table 3: **GMM estimation results of the standard neoclassical growth model**

IV Set	Small Sample Statistics	Parameters Estimates		Asymptotic St. Dev. Estimates**		J-statistic
		δ_T	γ^L_T	s_{δ_T}	$s_{\gamma^L_T}$	
Lag 0	Mean	0.990069	1.105563	0.000286	0.065376	4.490467
	Std. Dev.	(0.000460)	(0.084809)	(0.000161)	(0.048911)	(5.491713)
	$\sqrt{\text{MSE}}$		0.135411			
	Rejects* (%)					34.80
Lag 1	Mean	0.990080	1.147016	0.000275	0.055486	8.081440
	Std. Dev.	(0.000464)	(0.063156)	(0.000168)	(0.042841)	(7.364672)
	$\sqrt{\text{MSE}}$		0.160008			
	Rejects (%)					33.00
Lag 2	Mean	0.990085	1.174735	0.000265	0.047949	11.406011
	Std. Dev.	(0.000460)	(0.051896)	(0.000172)	(0.037244)	(8.923698)
	$\sqrt{\text{MSE}}$		0.182279			
	Rejects (%)					34.0
Lag 3	Mean	0.990092	1.191062	0.000256	0.042948	14.318159
	Std. Dev.	(0.000461)	(0.045439)	(0.000175)	(0.033666)	(10.286553)
	$\sqrt{\text{MSE}}$		0.196391			
	Rejects (%)					33.2
Lag 4	Mean	0.990098	1.203559	0.000249	0.039037	17.139470
	Std. Dev.	(0.000458)	(0.040682)	(0.000177)	(0.031593)	(11.909869)
	$\sqrt{\text{MSE}}$		0.207585			
	Rejects (%)					32.8
Lag 5	Mean	0.990105	1.212147	0.000243	0.036189	19.681924
	Std. Dev.	(0.000462)	(0.039145)	(0.000173)	(0.029025)	(13.651788)
	$\sqrt{\text{MSE}}$		0.215728			
	Rejects (%)					29.2

* I calculate the square root of the Mean Square Error of $\hat{\gamma}^L$, in the fourth column of the table. I also calculate the percentage of rejects in the J-statistic test at the 10% level, in the last column of the table.

** These are estimators of the diagonal elements of $\sqrt{\left(B'(S_w)^{-1}B\right)^{-1}/T}$.

Table 4: **Data Statistics for Model 1**

	$\lambda = 0$		$\lambda = 0.9$		$\lambda = 10.0$	
	T=10,000	T=115(500)	T=10,000	T=115(500)	T=10,000	T=115(500)
σ_u	0.0047	0.0047	0.0047	0.0047	0.0047	0.0047
σ_x	0.0414	0.0207	0.0416	0.0212	0.0626	0.0512
ρ_x	0.9937	0.9644	0.9831	0.9098	0.4303	0.1587
ρ_u	-0.0002	-0.0036	-0.0002	-0.0036	-0.0002	-0.0036
$\rho_{u,x}$	-0.0002	0.0020	0.1026	0.2251	0.7522	0.9098

Table 5: **Data Statistics for Models 2 and 3**

	$\lambda = 0$		$\lambda = 0.9$		$\lambda = 10.0$	
	T=10,000	T=115(500)	T=10,000	T=115(500)	T=10,000	T=115(500)
σ_u	0.0047	0.0047	0.0047	0.0047	0.0047	0.0047
σ_x	0.0414	0.0208	0.0417	0.0212	0.0629	0.0514
ρ_x	0.9937	0.9835	0.9831	0.9560	0.4300	0.2431
ρ_u	-0.0002	-0.0036	-0.0002	-0.0036	-0.0002	-0.0036
$\rho_{u,x}$	-0.0001	0.0021	0.1027	0.1633	0.7515	0.8627

Table 6: **Models with explanatory variables that are correlated with the residuals**

IV Set	Statistics	Model 1			Model 2	Model 3
		$\lambda = 0$	$\lambda = 0.9$	$\lambda = 10.0$	$\lambda = 0.9$	$\lambda = 0.9$
1	Bias	0.001367	0.001029	-0.003005	0.000020	0.000217 0
	Std. Dev.	0.019393	0.019259	0.026822	0.000349	0.003488
	Root MSE	0.019441	0.019287	0.026990	0.000349	0.003494
	% Rejects	-	-	-	-	-
	$(115)^{-1/2}\sigma_{\beta_T}$	0.016487	0.016409	0.019769	0.000337	0.003371
	$(115)^{-1/2}\sigma_{\beta}$	0.010506	0.010506	0.010487	0.000338	0.003383
2	Bias	0.001496	0.001749	0.004333	0.000024	0.000219
	Std. Dev.	0.019804	0.019825	0.017998	0.000353	0.003529
	Root MSE	0.019861	0.019902	0.018512	0.000354	0.003536
	% Rejects	11.4	11.4	7.8	11.4	11.6
	$(115)^{-1/2}\sigma_{\beta_T}$	0.016187	0.016106	0.017494	0.000331	0.003312
	$(115)^{-1/2}\sigma_{\beta}$	0.010503	0.010503	0.010483	0.000338	0.003382
3	Bias	0.001452	0.002592	0.011641	0.000024	0.000163
	Std. Dev.	0.019838	0.019938	0.017440	0.000360	0.003601
	Root MSE	0.019892	0.020106	0.020969	0.000361	0.003605
	% Rejects	11.2	11.4	4.2	11.6	11.6
	$(115)^{-1/2}\sigma_{\beta_T}$	0.015881	0.015788	0.015558	0.000324	0.003243
	$(115)^{-1/2}\sigma_{\beta}$	0.010501	.010501	0.010478	0.000338	0.003381
4	Bias	0.001194	0.003153	0.016765	0.000023	0.000111
	Std. Dev.	0.020491	0.020778	0.018955	0.000372	0.003723
	Root MSE	0.020526	0.021016	0.025305	0.000373	0.003725
	% Rejects	8.4	7.8	2.0	8.4	8.4
	$(115)^{-1/2}\sigma_{\beta_T}$	0.015597	0.015507	0.014827	0.000317	0.003170
	$(115)^{-1/2}\sigma_{\beta}$	0.010497	0.010475	0.010473	0.000338	0.003381
5	Bias	0.001500	0.000453	-0.006738	0.000019	0.000207
	Std. Dev.	0.020391	0.020267	0.388575	0.000349	0.003492
	Root MSE	0.020446	0.020272	0.388634	0.000350	0.003498
	% Rejects	-	-	-	-	-
	$(115)^{-1/2}\sigma_{\beta_T}$	0.017085	0.017518	1.540764	0.000338	0.003384 0
	$(115)^{-1/2}\sigma_{\beta}$	0.010573	0.010612	0.016024	0.000338	0.003383
6	Bias	0.001301	0.000767	0.008313	0.000022	0.000210
	Std. Dev.	0.020601	0.020545	0.066613	0.000355	0.003552
	Root MSE	0.020601	0.020560	0.067129	0.000356	0.003558
	% Rejects	12.4	10.0	13.8	9.2	9.2
	$(115)^{-1/2}\sigma_{\beta_T}$	0.017062	0.017313	0.046860	0.000337	0.003369
	$(115)^{-1/2}\sigma_{\beta}$	0.010572	0.010592	0.013593	0.000338	0.003383
7	Bias	0.001285	0.001430	0.017142	0.000014	0.000098
	Std. Dev.	0.021008	0.021260	0.029638	0.000365	0.003645
	Root MSE	0.021048	0.021308	0.034238	0.000365	0.003646
	% Rejects	8.6	10.0	14.2	8.8	8.8
	$(115)^{-1/2}\sigma_{\beta_T}$	0.016877	0.017155	0.018604	0.000331	0.003308
	$(115)^{-1/2}\sigma_{\beta}$	0.010567	0.010586	0.012228	0.000334	0.003353
8	Bias	0.000987	0.003862	0.021927	0.000015	-0.000025
	Std. Dev.	0.021004	0.021609	0.020181	0.000383	0.003834
	Root MSE	0.021028	0.021952	0.029801	0.000384	0.003834
	% Rejects	8.6	9.0	5.2	8.4	8.4
	$(115)^{-1/2}\sigma_{\beta_T}$	0.015168	0.015093	0.011747	0.000306	0.003066
	$(115)^{-1/2}\sigma_{\beta}$	0.010491	0.010469	0.010465	0.000334	0.003353

$(115)^{-1/2}\sigma_{\beta_T}$: Estimated asymptotic standard deviation of $\hat{\beta}$, where $\sqrt{(T)}(\hat{\beta} - \beta)$ converges in distribution to a $N(0, \sigma_{\beta})$ variable.

$(115)^{-1/2}\sigma_{\beta}$: Approximated true asymptotic standard deviation of $\hat{\beta}$ divided by square root of 115. (The value of σ_{β} has been calculated with 150 realizations of the stochastic process, of length 10,000).

Table 7: **Explicit Autoregressive models**

IV Set	1	2	3	4	5	6	7	8
				Model 4				
Bias	0.108367	0.062838	0.039063	0.027745	0.099558	0.055078	0.029764	0.027504
Std. Dev.	0.150245	0.178467	0.090399	0.078978	0.162301	0.100921	0.084154	0.074551
Root MSE	0.185248	0.189207	0.098478	0.083710	0.190404	0.114973	0.089262	0.079463
% Rejects	-	7.0	4.8	5.2	-	4.8	6.2	3.8
$(115)^{-1/2}\sigma_{\beta_T}$	31485.09	0.195115	0.107333	0.087238	44341.07	0.125225	0.089143	0.074324
$(115)^{-1/2}\sigma_{\beta}$	1.895718	1.155705	0.180338	0.130862	1.877793	0.175406	0.108483	0.108014
				Model 5				
Bias	-0.993760	-0.927624	-0.882428	-0.869678	-1.084149	-0.620163	-0.709384	-0.887656
Std. Dev.	7.443076	0.842647	0.237004	0.137479	7.710929	1.664495	0.447572	0.114031
Root MSE	7.509123	1.253212	0.913702	0.880477	7.786771	1.776273	0.838777	0.894950
% Rejects	-	0	0.8	1.2	-	0.6	2.0	0.8
$(115)^{-1/2}\sigma_{\beta_T}$	2845257035.9	6.666050	2.871392	1.974298	2043999121.9	13.911559	3.762227	1.508762
$(115)^{-1/2}\sigma_{\beta}$	938.75	10.515101	1.754676	1.240269	602.40	1.901722	1.082778	1.017307
				Model 6				
Bias	0.033300	-0.106472	-0.183367	-0.232481	0.034496	-0.022656	-0.121855	-0.280660
Std. Dev.	0.543472	0.347289	0.317907	0.301109	0.504009	0.425614	0.312009	0.293636
Root MSE	0.544491	0.363244	0.366999	0.380413	0.505188	0.426217	0.334960	0.406193
% Rejects	-	5.2	2.2	1.6	-	5.8	4.0	0.8
$(115)^{-1/2}\sigma_{\beta_T}$	1532084.7	0.310376	0.224202	0.195591	3022523.4	0.426661	0.277914	0.171504
$(115)^{-1/2}\sigma_{\beta}$	2.947335	1.573786	0.896816	0.742738	2.907589	0.944231	0.693756	0.655292

Table 8: **Model 7**

IV Set	1	2	3	4	5	6	7	8	9
				$\nu = 0.994$					
Bias	-0.942973	-0.986764	-1.010301	-0.967247	-1.011173	-1.005985	-0.890648	-0.935774	-0.973039
Std. Dev.	1.528205	0.609416	0.166664	1.519568	0.567583	0.176105	0.183553	0.117527	0.063638
Root MSE	1.795720	1.159781	1.023956	1.801292	1.159578	1.021283	0.909366	0.943125	0.975118
% Rejects	-	0.6	0.4	-	0.4	0.4	15.6	8.6	7.6
$(115)^{-1/2}\sigma_{\beta_T}$	19090792.2	2.312287	0.752616	40696416.7	4.770467	0.768655	0.614772	0.491206	0.266754
$(115)^{-1/2}\sigma_{\beta}$	84.284843	17.430160	6.484591	22.404810	12.342233	6.185077	1.679858	1.626872	1.525923
				$\nu = 0.194$					
Bias	-0.036266	0.053114	-0.039760	0.079916	-0.105413	-0.228255	0.030972	-0.033772	-0.091818
Std. Dev.	2.817204	0.208976	0.133310	2.824281	1.106984	0.306868	0.165420	0.109773	0.114227
Root MSE	2.817438	0.215620	0.139113	2.825412	1.111991	0.382450	0.168295	0.114850	0.146555
% Rejects	-	2.8	5.6	-	3.4	2.8	3.8	4.8	6.4
$(115)^{-1/2}\sigma_{\beta_T}$	182270879.4	0.633395	0.500891	234779412.5	2.416998	1.161621	0.565317	0.429534	0.422835
$(115)^{-1/2}\sigma_{\beta}$	127.735001	0.179347	0.137362	25.397377	0.505260	0.416081	0.153662	0.113644	0.113521

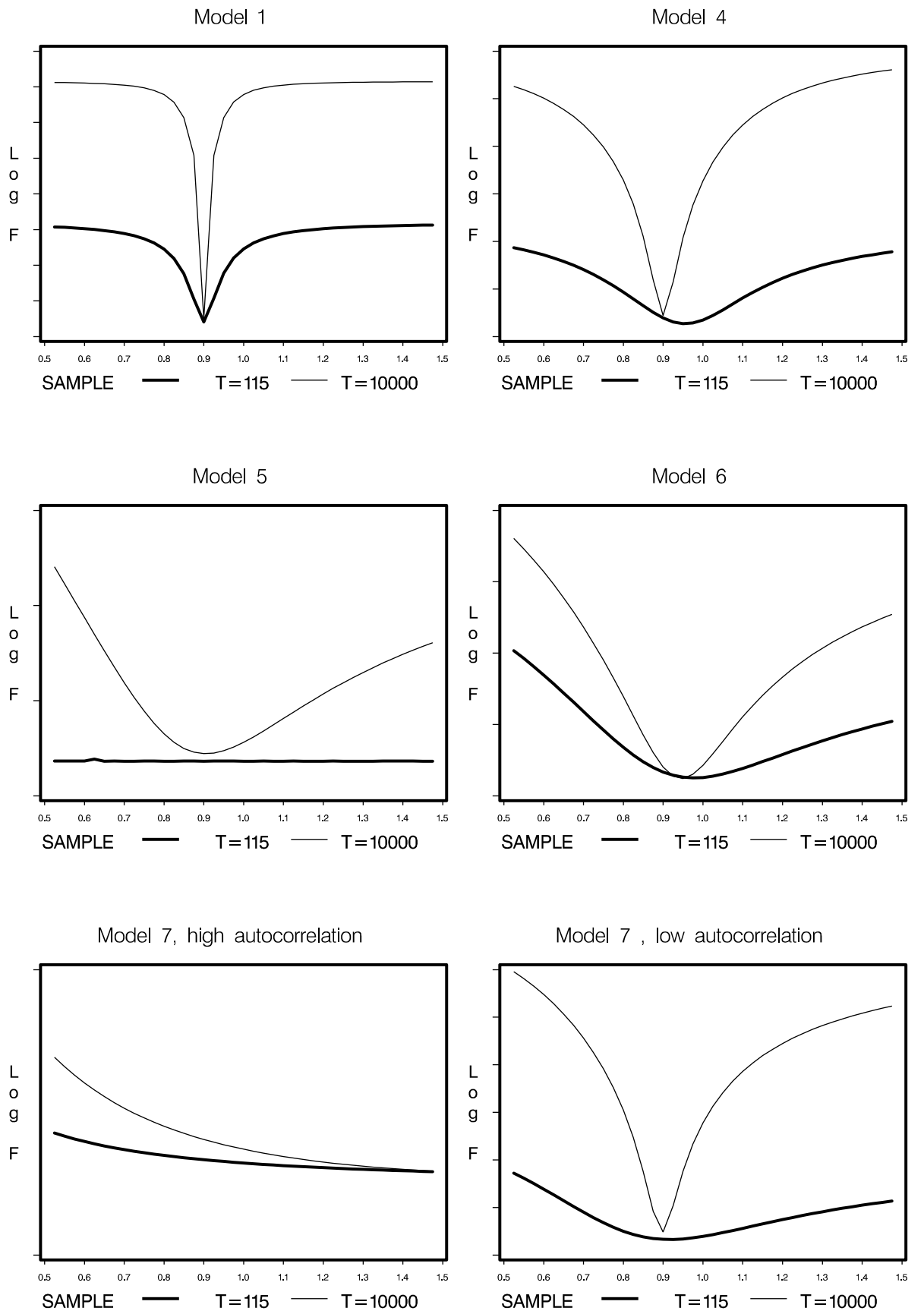# Figure 1: Objective Function: Instrument Set 7
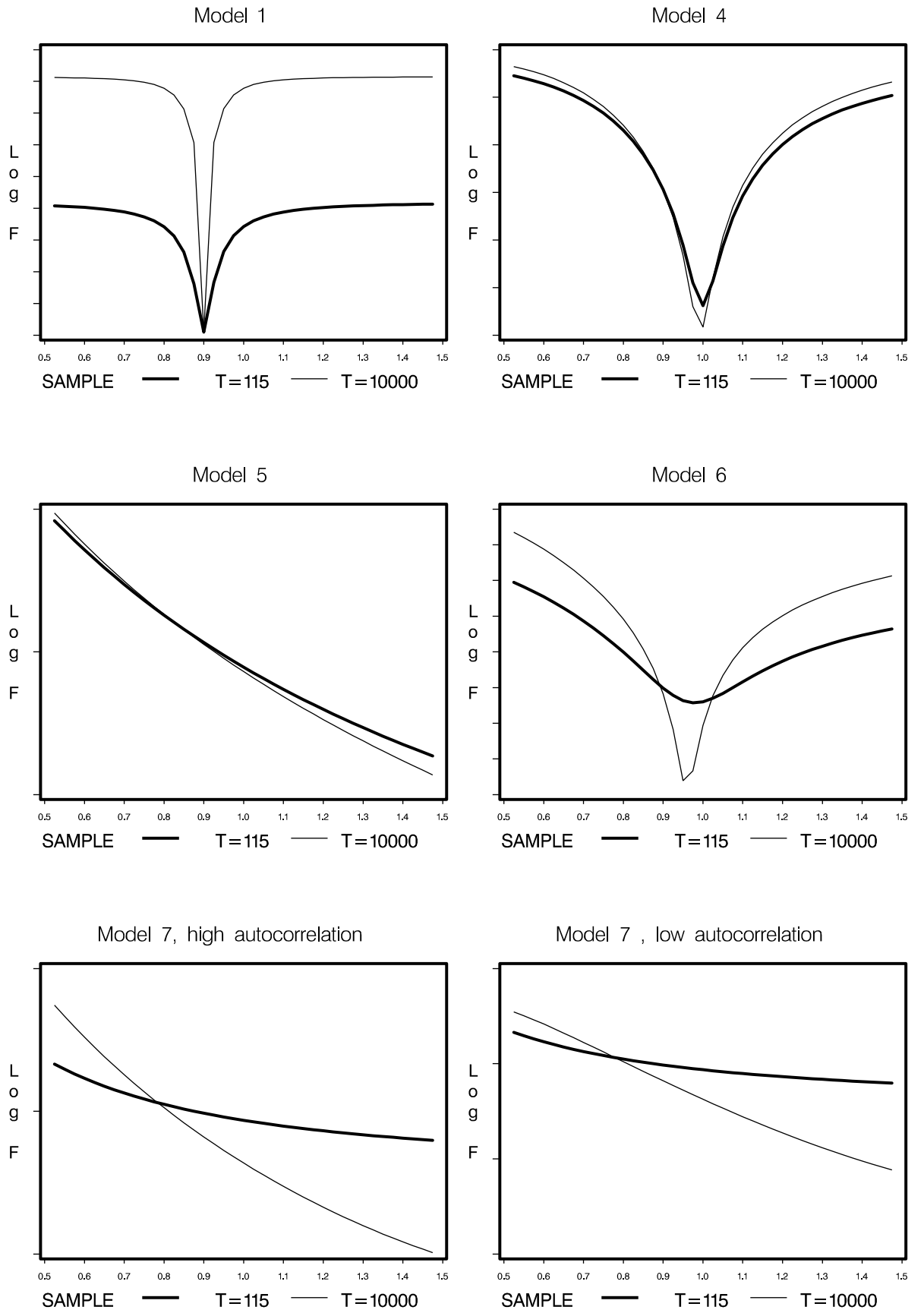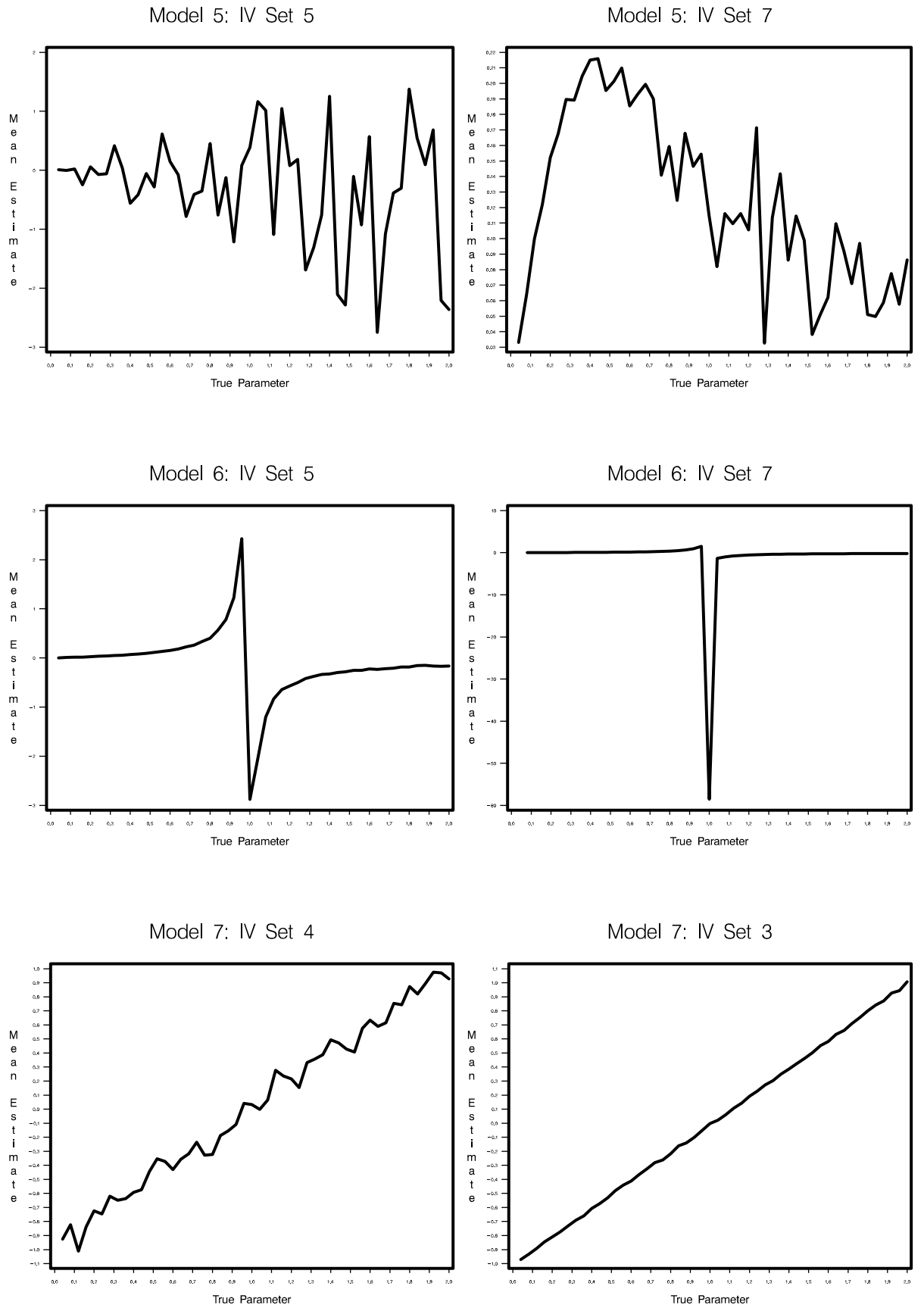

40

Figure 2: **Objective Function: Instrument Set 5**

41

Figure 3: **Relationship between the true parameter and the GMM estimate**

Model 5: IV Set 5

Model 5: IV Set 7

Model 6: IV Set 5

Model 6: IV Set 7

Model 7: IV Set 4

Model 7: IV Set 3

References

ALTONJI, J. G. AND L. M. SEGAL (94), "Small sample bias in GMM estimation of covariance structures, " unpublished manuscript, Northwestern University.

AMEMIYA, T. (66), "On the use of principal components of independent variables in two-stage least-squares estimation, " *International Economic Review*, 7(3):283–328.

ANDERSEN, T. G. AND B. E. SØRENSEN (95), "GMM estimation of a stochastic volatility model: a Monte Carlo study, " unpublished manuscript.

BURNSIDE, C. AND M. EICHENBAUM (94), "Small sample properties of generalized method of moments based Wald tests, " unpublished manuscript.

CHANG, P. H. AND K. L. JUDD (92). "Finite sample bias of generalized method of moments estimation, " Technical report, University of Southampton. Discussion Paper in Economics and Econometrics.

CHRISTIANO, L. J. AND W. DEN HAAN (95), "Small sample properties of GMM for business cycle analysis, " unpublished manuscript.

FUHRER J., G. MOORE AND S. SCHUH (93), "Estimating the linear-quadratic inventory model: Maximum likelihood versus generalized method of moments, " unpublished manuscript.

GARCIA-MILÀ T. , A. MARCET AND E. VENTURA (96). "Supply side interventions and redistribution, " Technical Report 115, UPF Working Paper Series.

GIOVANETTI G., A. MARCET AND R. MARIMON (93), "Growth, capital flows and enforcement constraints: The case of africa, " *European Economic Review*, 37:418–425.

HANSEN L.P, ,J.C. HEATON AND A. YARON (95), "Finite sample properties of some alternative GMM estimators, " unpublished manuscript.

HOROWITZ, J. AND P. HALL (96), "Bootstrap critical values for tests based on generalized-method-of-moments estimators, " *Econometrica*, 64:103–107.

KOCHERLAKOTA, N. R. (90), "Tests of representative consumer asset pricing models, " *Journal of Monetary Economics*, 26:285–304.

MARCET, A. AND W. J. DEN HAAN (94), "Accuracy in simulations, " *Review of Economic Studies*, 61(1)(206):3–18.

MARCET, A. AND D. A. MARSHALL (94a). "Convergence of approximate model solutions to rational expectation equilibria using the method of parameterized expectations, " Technical Report 17, UPF Working Paper Series.

——— (94b). "Solving nonlinear rational expectations models by parameterized expectations: Convergence to stationary solutions, " Technical Report 76, UPF Working Paper Series.

NELSON, C. R. AND R. STARTZ (90), "Some further results on the exact small sample properties of the instrumental variable estimator, " *Econometrica*, 58(4):967–976.

PAGAN, C. AND Y. JUNG (93), "Understanding some failures of instrumental variable estimators," unpublished manuscript.

R., M. AND A. MARCET (92), "Communication, commitment and growth," *Journal of Economic Theory*, 58(2).

ROGERSON, R. AND P. RUPERT (93), "On testing the intertemporal substitution of labor supply," *Journal of Economic Dynamics and Control*, 17:37–50.

STAIGER, D. AND J. H. STOCK (93), "Instrumental variables regression with weak instruments," unpublished manuscript.

TAUCHEN, G. (86), "Statistical properties of generalized method of moments estimators of structural parameters obtained from financial market data," *Journal of Business and Economic Statistics*, 4(4):397–416.

WEST, K. D. AND D. W. WILCOX (95), "A comparison of alternative instrumental variables estimators of a dynamic linear model," unpublished manuscript.