Cardiff Economics Working Papers

Working Paper No. E2022/10

Why does Indirect Inference estimation produce less small sample bias than maximum likelihood? A note

David Meenagh, Patrick Minford and Yongdeng Xu

May 2022

ISSN 1749-6010

Cardiff Business School
Cardiff University
Colum Drive
Cardiff CF10 3EU
United Kingdom
t: +44 (0)29 2087 4000
f: +44 (0)29 2087 4419
business.cardiff.ac.uk

Why does Indirect Inference estimation produce less small sample bias than maximum likelihood? A note

David Meenagh*
(Cardiff Business School, Cardiff University)

Patrick Minford
(Cardiff Business School, Cardiff University, and CEPR)

Yongdeng Xu
(Cardiff Business School, Cardiff University)

May 2022

Abstract

Maximum Likelihood (ML) shows both lower power and higher bias in small sample Monte Carlo experiments than Indirect Inference (II) and II's higher power comes from its use of the model-restricted distribution of the auxiliary model coefficients (Le et al. 2016). We show here that II's higher power causes it to have lower bias, because false parameter values are rejected more frequently under II; this greater rejection frequency is partly offset by a lower tendency for ML to choose unrejected false parameters as estimates, due again to its lower power allowing greater competition from rival unrejected parameter sets.

JEL Codes: C12, C32, C52

Keywords: Bias, Indirect Inference, Maximum Likelihood

*Corresponding Author: Cardiff Business School, Colum Drive, Cardiff, CF10 3EU, UK. Email: MeenaghD@cardiff.ac.uk

1 Introduction

In two recent surveys of indirect inference estimation Le et al. (2016) and Meenagh et al. (2019) have found by Monte Carlo experiment that in small samples the indirect inference (II) test has much greater power than direct inference in its most widely used form of maximum likelihood (ML). So much so that in practice the power of the II procedure needs to be limited by reducing the size of the auxiliary model in order to ensure finding a tractable model that can pass the test threshold. These surveys also found that in small sample estimation II produced much lower bias than ML. Meenagh et al. (2019) noted (p.606): 'This property (of low small sample bias) comes from the high power of the test in rejecting false parameter values.' In this note we attempt to quantify this small sample relationship between power and bias under ML and II.

Let us first recap each procedure. In ML the structural model is taken to the data and the estimation searches over its parameters, including those of the ARMA error processes, to minimise the sum of squared reduced form residuals. The joint likelihood of the data, conditional on the model, is maximised when this sum is minimised.

By contrast, in II, the data is first represented by an auxiliary model, which is simply a set of parameters which best describe the data behaviour. For expositional purposes we will take these to be the parameters of the unknown structural model's reduced form VAR — as Meenagh et al. (2019) demonstrate, II gives approximately the same results whatever the form of the auxiliary model, provided each form has the same number of descriptive parameters — the forms they explore are moments and impulse response functions. Suppose we examine the VAR parameters, we can think of the structural model we are estimating as implying a joint normal distribution of these reduced form parameters, which we illustrate for two parameters in Figure 1.

We can generate this Likelihood distribution of the two parameters, ρ_1,ρ_2, by bootstrapping the structural model with its shocks and estimating a VAR on each bootstrap. The cumulative probability of these two parameters' squared deviation from the model's mean prediction (the peak likelihood point) weighted by the inverse of their variance-covariance matrix, V, is represented by a chi-squared distribution where k, the degrees of freedom, is given by the number of VAR parameters. If the two parameters have a low correlation, then each is weighted by 1/its variance. The weight on ρ_1 falls relative to the other's with a rising covariance/its variance.

In Figure 1 one can see the likelihood distribution of the different data-estimated reduced form coefficients, ρ_1,ρ_2, according to the model parameters- the top frame showing one with zero correlation between the two ρs , the bottom frame one with a high positive correlation. In II the parameters of the model are searched over to find those that have the highest likelihood, given the data-estimated coefficients shown by the red or blue dots; the parameters whose peak likelihood gets closest to the data dots will be the II estimates. In ML the red or blue dots of the data are directly taken as the ML reduced form coefficients; and the model structural parameters are reverse-engineered to produce the ones closest to them.

Thus take a model $y = f(\theta; \epsilon)$ which has a reduced form $y = v(\theta; \epsilon)$. Assume it is identified so that there is a unique v corresponding to a particular f; thus given v we can discover f and vice versa. Suppose now on a sample y_0 we obtain an estimate $\widehat{v}(y_0)$. In II we compute the likelihood of $\widehat{v}(y_0)$ conditional on the model and the data, thus $L[\widehat{v}(y_0) \mid$

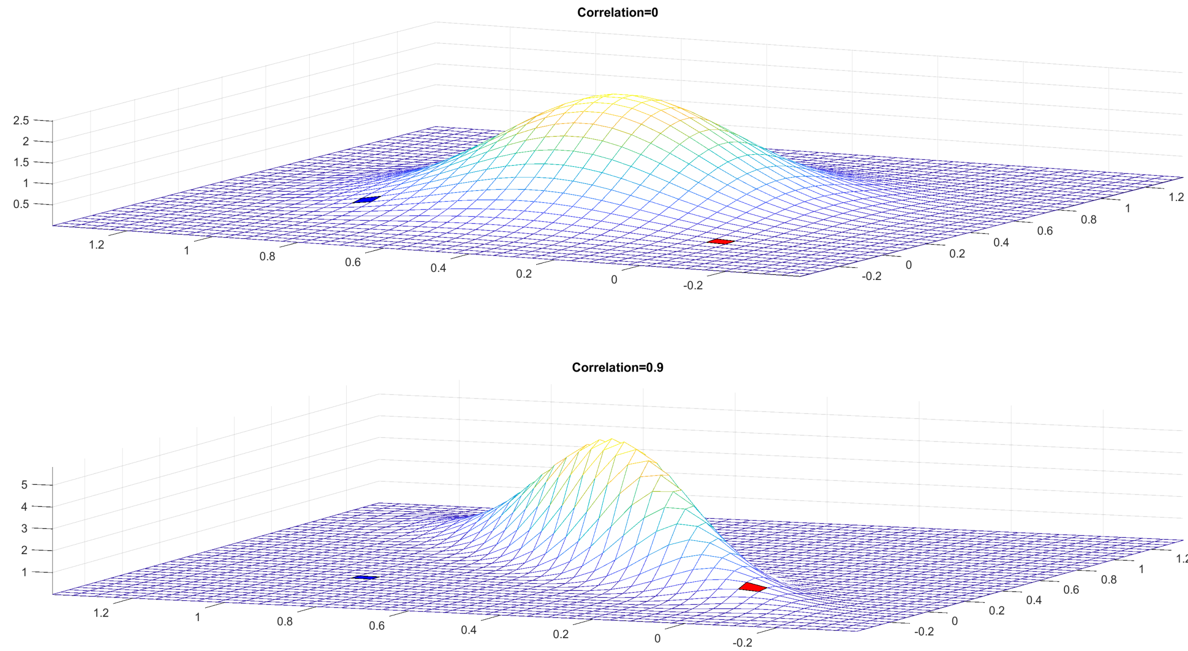

Figure 1: Bivariate normal distribution with correlation of 0 and 0.9. Two possible data points shown: x=0.1, 0.9 and y=0.0

$y_0, f(\theta;\epsilon)]$; we then search over θ to find the maximum likelihood; this is the II estimate. If unbiased, it will on average be the f corresponding to v. In general we find low bias in II. In terms of our diagram $\widehat{v}$ is the blue or red dot and the joint distribution of the estimated model will be close to being centred around it. Now ML in principle does the same, choosing the ML values of θ that generate $\widehat{v}$ as their solution of $y_0 = f(\theta;\epsilon)$.

It would seem therefore that the two estimates of the structural parameters should be the same. Indeed, it has been shown (e.g. by Gourieroux et al, 1993) that this is the case asymptotically, i.e. for very large samples. Both estimators are consistent in large samples, implying no bias.

However, in small samples — such as are typical in macroeconomics — they are not typically the same and we find bias in both according to our Monte Carlo experiments.

2 Explanation for the difference in small sample bias

The question we wish to answer here is why the two estimates differ in small samples and the quantitative contribution of the causes.

Le et al. (2016) showed that the power of the II and the ML-based LR tests of the model $f(\theta;\epsilon)$ differed; specifically II was substantially more powerful. This occurred when the II test used as the distribution of v implied by $f(\theta;\epsilon)$ the model-restricted distribution. If on the other hand it used the distribution of v from the reduced form data-implied distribution, then the power of II was reduced to equality with that of LR. Thus the power of the II test was considerably greater than that of the ML-based LR test — the reason being that the II test used the distribution of v as restricted by the model under test, whereas the LR test used the reduced form v distribution from the data. In Figure 2 we show a stylised illustration of this point: the figure shows the situation for the likelihood distribution of $\widehat{v}(\phi)$,the vector of auxiliary model features (ordered according to their Wald value under the model, with parameter vector ϕ, indicated), under the restricted and unrestricted cases. To the left we see the distribution under the true model, with ϕ_{TRUE}; to the right we see the distribution under the false model, ϕ_{FALSE}. In the top panel this is given by the unrestricted distribution taken from the data, which is the same as the left hand distribution. In the bottom panel, it is given by the distribution generated by the false parameter model in conjunction with the errors implied by the model and the data. It can be seen that this latter distribution lies more narrowly around the central false average due to the inward pull of the false parameters on the simulations.

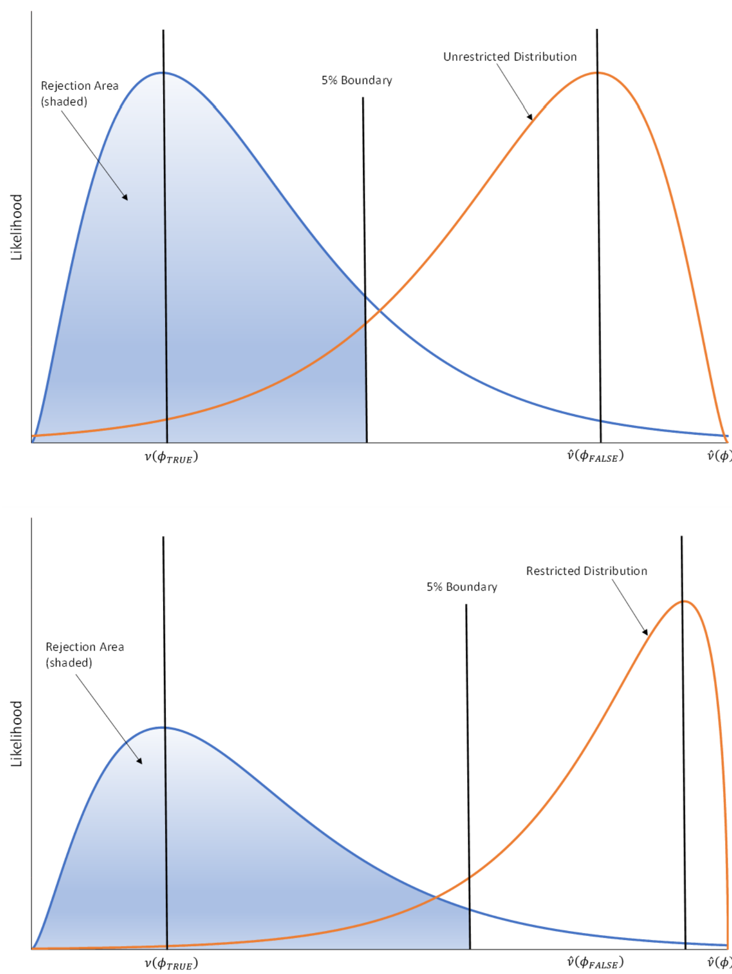

Figure 2: Comparison of rejection rates of unrestricted and restricted distributions of $\widehat{v}(\phi_{FALSE})$

Table 1 shows the relative power of the II and ML tests on a 3 variable VAR (1) and is replicated from Le et al. (2016) Table 1, where the Direct Inference column shows the results based on the LR test.

Percent Mis-specified	Indirect Inference	Direct Inference
True	5.0	5.0
1	19.8	6.3
3	52.1	8.8
5	87.3	13.1
7	99.4	21.6
10	100.0	53.4
15	100.0	99.3
20	100.0	99.7

Table 1: Rejection Rates for Wald and Likelihood Ratio for 3 Variable VAR(1)

We can now turn to the implications of this greater power in II testing for the bias that arises in estimation by II and ML on small samples. The bias we estimate in our Monte Carlo (MC) experiments is defined as $B = E(\widehat{\theta}) - \theta$, where the expectation is across all the MC pseudo-samples from the true model. We can express this definition in terms of all the possible sets of θ arranged in order of falsity, thus $B = \Sigma_{i=\%F}(\theta_i - \theta)P_i$ where each θ_i is the set of parameters of $i\%$ falseness and P_i is the frequency with which these are estimated in the MC samples. We can think of estimation by II or ML as a process related to rejecting model parameters that fail each test respectively; if a false parameter set is rejected, it cannot become an estimate, and if not rejected for a sample, it can go on to become an estimate for that sample. We also need to know the probability for either II or ML that, conditional on not being rejected, a parameter set θ will then be chosen as an estimate. Call these probabilities in turn P_1 for the probability of non-rejection, and P_2 for the probability of selection conditional on non-rejection. The MC experiments give us directly P_{i1} as one minus the rejection rate for θ_i.while we can obtain P_i from our MC results directly as the proportion of estimates that are False to each extent. Then we derive P_{i2} from $P_i = P_{i1} \times P_{i2}$. To gauge P_{i2} we argue as follows: a θ parameter set that has not been rejected will still not be selected as an estimate if there is an unrejected θ of lesser falseness available instead that dominates it in the competition to become an estimate.

Table 2 shows the small sample bias of the two estimators in the Monte Carlo experiment, replicated from Table 3 from Le et al. (2016), clearly showing the big reduction in the bias under II versus ML.

We show next the predicted two probabilities and biases for II and ML in Table 3. For this table we have repeated the bias analysis with a fresh set of 1000 samples from the same model, yielding different absolute mean biases, as one would expect; in this set the ML bias is about the same, the II bias rather smaller. What we see is that on average an unrejected θ is 60% more likely to survive to being estimated under II as under ML. We suggest this is because II has a generally higher rejection rate than ML, so that an unrejected θ faces less competition from other unrejected θ, and so has a greater probability of surviving to estimation. Under ML the probability of survival is inversely correlated with the probability

		Starting (true) coef	Mean Bias (%) II	Mean Bias (%) FIML	Absolute Mean Bias (%) II	Absolute Mean Bias (%) FIML	Ratio II/FIML
Steady-state elasticity of capital adjustment	φ	5.74	−0.900	5.297	0.900	5.297	0.16
Elasticity of consumption	σ_c	1.38	−5.804	−7.941	5.804	7.941	0.73
External habit formation	λ	0.71	−13.403	−21.240	13.403	21.240	0.63
Probability of not changing wages	ξ_w	0.70	−0.480	−3.671	0.480	3.671	0.13
Elasticity of labour supply	σ_L	1.83	0.759	−8.086	0.759	8.086	0.093
Probability of not changing prices	ξ_p	0.66	−1.776	0.027	1.776	0.027	65.8
Wage indexation	ι_w	0.58	−0.978	6.188	0.978	6.188	0.158
Price indexation	ι_p	0.24	0.483	3.228	0.483	3.228	0.15
Elasticity of capital utilisation	ψ	0.54	−13.056	−29.562	13.056	29.562	0.44
Share of fixed costs in production (+1)	Φ	1.50	−1.590	2.069	1.590	2.069	0.75
Taylor Rule response to inflation	r_p	2.04	7.820	2.815	7.820	2.815	2.78
Interest rate smoothing	ρ	0.81	−0.843	−0.089	0.843	0.089	9.47
Taylor Rule response to output	r_y	0.08	−4.686	−29.825	4.686	29.825	0.16
Taylor Rule response to change in output	$r_{\Delta y}$	0.22	−5.587	0.171	5.587	0.171	32.7
Average			−2.861	−5.758	4.155	8.586	0.48

Table 2: Bias of Indirect Inference and FIML

of non-rejection of the neighbouring θ closer to the truth: we suggest this is because the higher the chances of their non-rejection, the greater is the competition from them — see the bottom frame of Figure 3. What we see under II is different — the top frame of Figure 3. Survival chances of false θ, if unrejected, are low at the two extremes — both when close to true and when extremely false. Thus competition from better alternatives is greatest either close to the truth (when the truth is a serious rival), or very far from the truth (when the less absurdly false are serious rivals). This shift of survival probability to the extremes weakens the tendency for II to reduce bias, by increasing the estimation chances of the middlingly false values which contribute most to the bias after taking account of rejection.

$\theta : \%False - II$	P_{i1}	P_{i2}	P_i	$\theta : \%False - ML$	P_{i1}	P_{i2}	P_i
1	0.80	0.09	0.07	1	0.94	0.00	0.00
2	0.64	0.81	0.52	2	0.92	0.02	0.02
3	0.48	0.61	0.29	3	0.91	0.08	0.06
4	0.31	0.27	0.08	4	0.89	0.02	0.016
5	0.13	0.13	0.02	5	0.87	0.02	0.018
6	0.07	0.07	0.01	6	0.82	0.05	0.042
7	0.01	0.00	0.00	7	0.78	0.16	0.122
				$8-9$	0.62	0.54	0.332
10	0.00		0.00	10	0.47	0.43	0.20
				$11-14$	0.40	0.58	0.23
15	0.00		0.00	15	0.00		0.00
20	0.00		0.00	20	0.00		0.00
Predicted Bias* $= E(\widehat{\theta}) - \theta$	5.9	0.42	2.46	Predicted Bias* $= E(\widehat{\theta}) - \theta$	33	0.26	8.7

*The entries for this row, for each of II and ML, are in turn: $\sum_i P_{i1}\theta_i$;$[\sum_i P_i\theta_i]/[\sum_i P_{i1}\theta_{i,}]$; and $\sum_i P_i\theta_i$

Table 3: Predicted probabilities and bias

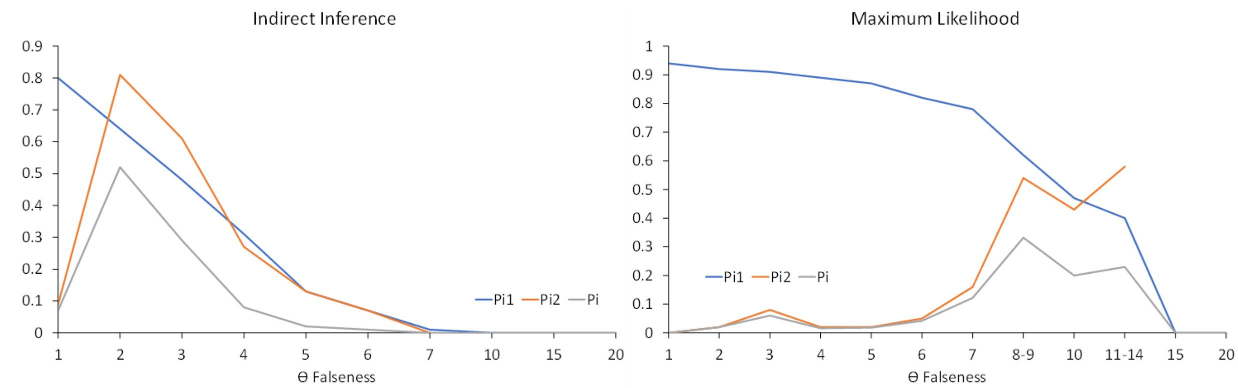

Figure 3: Predicted probabilities and biases for II and ML

Summarising our findings, our Monte Carlo experiments have shown that the lower bias of II compared to ML comes primarily from a much higher rejection rate of false coefficients. This advantage is to a modest extent offset by the higher probability under II that unrejected false coefficients will survive to become estimates. We interpret this in terms of the competition between unrejected coefficients: this is greater under ML than II because there are more unrejected coefficients to choose from at all levels of falseness. This competition also behaves differently across the range of falseness, increasing with falseness under ML as nonrejection falls, but intensifying under II at both extremes, either close to truth or highly false.

3 Conclusions

In this note, we have reflected on the reasons that Maximum Likelihood (ML) shows both lower power and higher bias in small samples than Indirect Inference (II), drawing on the earlier work of Le et al. (2016) and Meenagh et al. (2019), based on extensive Monte Carlo experiments. It emerges from this work that when ML is being used, the likelihood distribution of $\widehat{v}$, the auxiliary parameter vector from the model under test, has a variance given by the unrestricted distribution of the errors whereas when II is used it is given by the variance of their distribution as restricted by the θ of the model being tested, which is much smaller. This is the source of the higher power of II, as explained by Le et al. (2016). This in turn implies that II will have lower bias, because as sample data from the true model varies, false parameter values will be rejected much more frequently under II; this greater rejection frequency is partly offset by a lower tendency for ML to choose unrejected false parameters as estimates, due again to its lower power allowing greater competition from rival unrejected parameter sets.

References

[1] Gourieroux, C., Monfort, A., Renault, E. (1993). 'Indirect inference'. *Journal of Applied Econometrics* 8, Supplement: Special Issue on Econometric Inference Using Simulation Techniques, S85–S118.

[2] Le, V.P.M., Meenagh, D., Minford, P., Wickens, M. and Xu, Y. (2016) 'Testing Macro Models by Indirect Inference: A Survey for Users',*Open Economies Review*, 2016, 27(1), 1-38.

[3] Meenagh, D., Minford, P., Wickens, M. and Xu, Y. (2019) 'Testing Macro Models by Indirect Inference: A Survey of recent findings', *Open Economies Review*, 2019, vol. 30, issue 3, No 8, 593-620.