Why is estimating AB models so weird?

Matteo Richiardi [a b]
Lisa Sella [c]
Jakob Grazzini [a]

[a]University of Turin
[b]LABORatorio R. Revelli Center for employment studies
[c]CERIS C.n.r.

GSDP AB modelling workshop, Sorbonne, Paris, September 8-10, 2011

Motivation

- A common critique towards agent-based models (ABMs) is that they often remain at a theoretical level, and are not "taken to the data"
- Only few examples exist on structural estimation of ABMs: very simple models, few parameters, little is known about the properties of the estimates
- *What's so weird about estimation of ABMs? Is it just computational complexity?*

NO! It's computational complexity + inherent characteristics of ABMs (interaction (with long-lasting effects), nonlinearities, focus on + out-of-equilibrium dynamics)

Running examples: Models of innovation diffusion

1. Simple statistical model (purely descriptive): individuals adopt at a rate

$$h(t) = p\lambda t^{p-1}$$

   (Weibull hazard rate)

2. Bass[1969] new product adoption model: the rate of adoption at time t is a function of an *external* force (marketing effort) and an *internal* one (word-of-mouth)

$$h(t) = \frac{f(t)}{S(t)} = p + qF(t)$$

3. AB version of Bass (ABBass): global interaction is replaced by local interaction, given some specific network structure

$$h(t) = \frac{f(t)}{S(t)} = p + qA(t)$$

All lead to the characteristic S-shaped cumulative adoption curve [Geroski, 2000].

Outline of the presentation

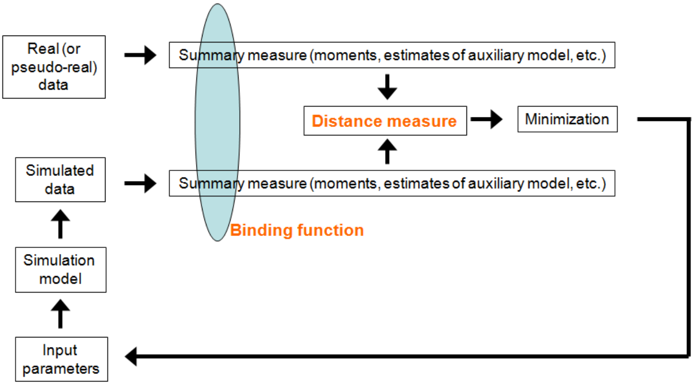
- We first review some econometric methods that can be used for estimating ABMs (focus on MSM in particular)
- We then show by means of Montecarlo analysis how their nice properties are lost when interaction (with long-lasting effects) is introduced. This is responsible for the biasedness and inconsistency of Bass estimates on early periods of the diffusion process. Considering local interaction makes the problem worse, as the estimates are biased and inconsistent even when the whole process is observed.

Existing attempts at estimating ABMs

- [Winker and Gilli, 2001; Gilli and Winker, 2003] estimate by MSM respectively 2 and 3 parameters of an ABM of the foreign exchange market based on [Kirman, 1991, 1993]: focus on optimization heuristics
- [Winker et al., 2007] deal with the problem of moments selection, and propose a set of statistics on exchange rate returs to estimate models of exchange rate
- [Boswick et al., 2007] estimate by NLLS a dynamic asset pricing model characterized by agents with heterogeneous beliefs
- [Alfarano et al., 2005, 2006] estimate agent-based models that are simple enough to derive a closed form solution for the distribution of relevant statistics

Indirect estimation

Simulation based techniques [Gouriereux and Monfort, 1996; Stern, 1997, 2000; Smith, 2008]. The basic idea is to replace the evaluation of an analytical expression that is impossibile or too complicated to compute with its numerical counterpart: for large samples the numerical counterpart tends to the analytical expression, and can then be confronted with its real sample value in the estimation procedure.



Simulated minimum distance estimators

Remarks

- The expression has to be computed just once in the real data (which does not change), and once every iteration until convergence in the artificial data, as its value depends on the the structural parameters. The change in the value of the parameters of each iteration is determined according to some optimization algorithm, with the aim to minimize the distance.
- Minimization requires that the same series of random draws is used for each iteration of the simulated model, in order to insulate from the stochastic component of the model. Lacking this, the minimization algorithm might well get stucked in cycles.
- In order for the estimates to have desirable properties as *unbiasedness, consistency, efficiency, asymptotic normality, etc.* convenient statistics have to be chosen. These statistics ot only have to be theoretically appropriate but also empirically relevant, i.e. sensitive to the value of the structural parameters, θ.

Simulated minimum distance estimators

The path estimator

Given that the output of interest is a time series, a natural criteria would be to compare the two paths y_t (real) and $y_t^s(\theta)$ (simulated). This however does not guarantee consistency.

To see why, suppose the real world DGP is $y_t = r(u_t, \theta^*)$, where the u_t are i.i.d. random variables with known p.d.f. $f(u)$.
The path estimator is (assuming for simplicity $T^s = T$):

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{T} (y_t^s(\theta) - y_i)^2$$

Under usual regularity conditions, the estimator tends asymptotically to the solution $\hat{\theta}_\infty$ of the limit problem.

Consistency requires $\hat{\theta}_\infty = \theta^*$, a condition that is not satisfied in general. For instance, suppose that u_t is exponential, with $f(u) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}$, for $x > 0$. Then $E(x) = \theta$, $var(x) = \theta^2$, and

$$\hat{\theta}_\infty = \arg\min_{\theta} [\theta^2 + \theta^{*2} + (\theta - \theta^*)^2] \Leftrightarrow \hat{\theta}_\infty = \theta^*/2 \neq \theta^*$$

This example illustrates the difference between estimation and calibration. Comparing the real and simulated paths is a convenient way to calibrate a model, but not to estimate it.

Simulated minimum distance estimators

Different approaches

In the literature a number of different simulation-based estimation techniques have been proposed. Out of these, three approaches have been singled out [Billio, 2002]:

- A first approach relies on simulation based methods which are relatively simple to implement, at the cost of achieving sub-optimal efficiency: the Method of Simulated Moments (MSM) [McFadden, 1989; Pakes, 1989; Lee and Ingram, 1991; Duffie and Singleton, 1993], the Indirect Inference Method (II) [Smith, 1993; Gourieroux et al., 1993], the Pseudo-Maximum Simulated Likelihood Method [Laroque and Salanié, 1993], the Efficient Method of Moments [Gallant, 1996].
- A second approach aims at approximating the likelihood function through importance sampling methods [Danielsson and Richard, 1993] or Markov Chain Montecarlo techniques [Durbin and Koopman, 1997].
- A third approach is based on Bayesian methods.

Here, we review only the simpler methods, which appear to be the most suited for applications to ABMs.

The method of simulated moments

Immediately after considering the actual paths, a second natural choice for comparing real and simulated output is to measure the difference between some empirical moments computed on y_t and on $y_t^s(\theta)$.

In the example above, the (first) moment estimator is:

$$\hat{\theta} = \arg\min_{\theta} \left[\frac{1}{T} \sum_{t=1}^{T} y_i - \frac{1}{T} \sum_{t=1}^{T} y_i^s(\theta) \right]^2$$

If we denote the first moment as:

$$E[y_t] = \int r(u, \theta^*) f(u) du = k(\theta^*)$$

the moment estimator converges to the solution $\hat{\theta}_\infty$ of the limit problem:

$$\hat{\theta}_\infty = \arg\min_{\theta} [E(y_i) - E(y_i^s(\theta)]^2 = \arg\min_{\theta} [k(\theta^*) - k(\theta)]^2 \quad (1)$$

which is $\hat{\theta}_\infty = \theta^*$. The estimator $\hat{\theta}$ is therefore consistent.

The method of simulated moments

Weighting

- More generally in the MSM (as in the simulated general method of moments) different order of moments of the time series of interest are used, and then weighted to take into account their uncertainty.
- The intuition is to allow parameters estimated with a higher degree of uncertainty to count less, in the final measure of distance between the real and artificial data [Winker et al., 2007]. Having different weights (or no weights at all) impinges on the efficiency of the estimates, not on their consistency.
- Note that while the uncertainty regarding the simulated moments can be reduced by increasing the number of simulation runs and the size of the simulated population, the uncertainty in the estimation of the real moments on the basis of real sample data cannot be reduced.
- If the number of moments is equal to the number of structural parameters to be estimated, the model is just-identified. The minimized distance, for the estimated values of the parameters, is therefore 0 in the limit (as the sample size grows bigger), supposing the model is correctly specified. If the number of moments is higher than the number of parameters the model is over-identified and the minimized distance is always positive. If it is lower it is under-identified.

The method of simulated moments

Example: The Weibull model

Let's suppose the *rwDGP* is such that individuals exit the state, at time T_i, at a rate

$$h(t) = p\lambda t^{p-1}$$

This is a Weibull duration model, with λ as a scale factor.

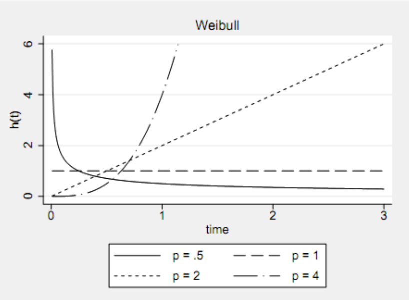


Figure: Example plot of Weibull hazard functions

Let's assume $p^* = 2$ and that we observe all durations.
Remark: nothing changes if only an early phase of the process is observed, and some (most) spells are right-censored.

The method of simulated moments

Example (cont'd)

Suppose that we have a (well specified) model for this process, and we want to estimate the parameter p. The theoretical mean time to failure (MTTF) is

$$E[T_i^{rw}] = \frac{1}{\lambda^{1/p}} \Gamma \left(1 + \frac{1}{p}\right)$$

This is a moment condition. As an estimate for $E[T_i^{rw}]$ we take the average observed time to failure, $\bar{T}_i$. Since in general the expression for the theoretical mean on the r.h.s. is not known, or it cannot be inverted in order to get an estimate for p, the MSM prescribes to simulate it. Hence, the moment condition becomes

$$E[\bar{T}_i - \bar{T}_i^s(p)] = 0$$

which implies

$$\hat{p} = argmin(\bar{T}_i - \bar{T}_i^s(p))^2$$

The method of simulated moments

Example (cont'd)

However, the choice of the MTTF as the binding function is not a good one. The reason is that, with positive duration dependence, more than one value of the parameter p can lead to the same MTTF.

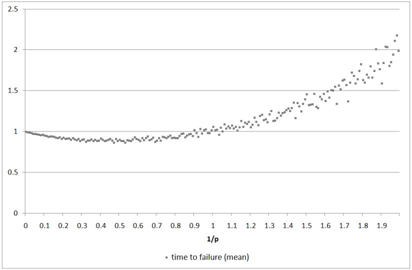


Figure: Mean time to failure for Weibull model, $\lambda = 1$. Simulation over 1,000 individuals.

The intuition is that very high values of p (low values of $1/p$ in the graph) imply a very small exit probability for small durations; however, the hazard rate increases quickly, so that very long durations are also very unlikely. On the other hand, somewhat lower values of p imply a higher probability of observing very short durations, but also a higher probability of observing very long durations. The two effects, for given values of p, counterbalance perfectly.

The method of simulated moments
Example (cont'd)

The second order moment, on the other hand, is monotonic in p, and is used in the estimation procedure:

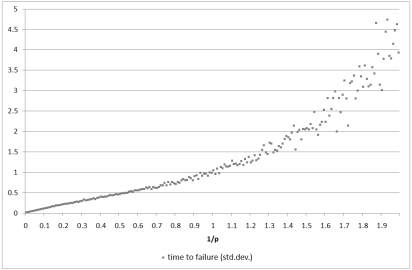


Figure: Standard deviation of time to failure for Weibull model, $\lambda = 1$. Simulation over 1,000 individuals.

The method of simulated moments

Example (cont'd)

The estimation procedure is tested by means of a Montecarlo experiment (100 replications).

Population size	10	50	100	500	1,000	5,000	10,000
Mean of estimated coeff.	0.543	0.507	0.500	0.499	0.498	0.499	0.499
Var of estimated coeff.	0.03870	0.00919	0.00397	0.00071	0.00034	0.00007	0.00003

Table: Montecarlo results for MSM estimation of the Weibull model with parameter $1/p = .5$.

The fact that the estimates are always centered around the true value 0.5 shows that the estimation procedure is unbiased. Consistency is shown by the reduction in the variability of the estimated coefficients, as the sample size increases: when the population size doubles, the variance halves.

The distribution of the estimator is asymptotically normal. Hence, confidence intervals can be constructed from the stadard deviation of the (bootstrapped) estimated coefficients.

Indirect Inference

In the II method, the basic idea is to use the coefficients of an auxiliary model, estimated both on the real and on the simulated data, to describe the data, *i.e.* as summary statistics on the original model. Hence, the method prescribes the following steps:

1. simulate the model for given values θ_0 of the parameters and obtain artifical data;
2. estimate the parameters of an auxiliary model;
3. change the structural parameters of the original model until the distance between the estimates of the auxiliary model using real and artifical data is minimized.

The auxiliary model can be overly simple and misspecified; however, the estimates are more efficient if it is a good statistical description of the data, i.e. a 'bona fide' *reduced form* version of the model. The properties of II method, on the other hand, crucially depend on a correct specification of the structural model (some semiparametric methods have been proposed that make II more robust to the structural model specification [Dridi and Renault, 2000]).

Indirect Inference
(cont'd)

Let $y_t^s = m(X, \theta)$ be the *model DGP*, with X explanatory variables (possibily including lagged endogenous variables) and θ parameters, and let $y_t = a(Z, \beta)$ be an *auxiliary model*, with Z explanatory variables (possibily including lagged endogenous variables) and β parameters.
$\hat{\beta}$ are the estimates of β using real data y_t, while $\tilde{\beta}(\theta)$ are the estimates of β using artificial data y_t^s, with parameters θ.

The II estimator is:

$$\hat{\theta} = \arg\min_{\theta} \left[\hat{\beta} - \tilde{\beta}(\theta)\right]' \Omega^{-1} \left[\hat{\beta} - \tilde{\beta}(\theta)\right] \quad (2)$$

where Ω is a positive definite matrix.

As in the method of simulated moments, if the number of the parameters of the auxiliary model is equal to the number of parameters in the original model the original model is just-identified, and the distance between the estimated coefficients on the real and on the simulated data, if the model is correctly specified, goes in the limit to zero. If the number of parameters in the auxiliary model is bigger than the number of parameters in the original model, the original model is over-identified, and the distance between the estimated coefficients remain positive. If the number of parameters in the auxiliary model is smaller than the number of parameters in the original model, the original model is under-identified.

Indirect Inference

Example

In the Weibull example presented above, it could be tempting to opt for a very simple auxiliary model in the form of an exponential model. Exponential models are particular cases of the Weibull models, with $p = 1$. This implies a constant hazard rate. Inference in this case is theoretically possible by comparing the scaling factor λ in the pseudo-true and in the simulated data. However, a Montecarlo experiment similar to the one already described shows that the exponential model is too poor a description of the Weibull model, for $p = 2$:

Population size	10	50	100	500	1,000	5,000	10,000
Mean of estimated coeff.	0.298	0.421	0.436	0.489	0.463	0.475	0.459
Var of estimated coeff.	0.07344	0.05805	0.05874	0.02262	0.01596	0.00877	0.00508

Table: Montecarlo results for II estimation of the Weibull model with parameter $1/p = .5$. The auxiliary model is exponential.

Indirect Inference
Example (cont'd)

The log-logistic specification appears to be a better choice for our auxiliary model. In the log-logistic model, the hazard is

$$h(t) = \frac{\lambda^{1/\gamma} t^{1/\gamma - 1}}{\gamma[1 + (\lambda t)^{1/\gamma}]^2}$$

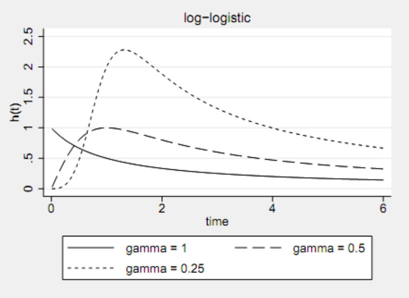


Figure: Example plot of Log-logistic hazard functions

As a proxy for the Weibull, it immediately appears more apt than the exponential.

Indirect Inference

Example (cont'd)

The Montecarlo experiment confirms that the choice is correct: although the log-logistic model is misspecified, the estimated parameters for p are centered around the true value, with a variance that declines at the usual rate with sample size.

Population size	10	50	100	500	1,000	5,000	10,000
Mean of estimated coeff.	0.523	0.497	0.495	0.502	0.499	0.499	0.500
Var of estimated coeff.	0.03901	0.00627	0.00367	0.00071	0.00038	0.00009	0.00004

Table: Montecarlo results for II estimation of the Weibull model with parameter $1/p = .5$. The auxiliary model is log-logistic.

Note that the efficiency of the MSM estimates and the II estimates, as measured by the variance of the estimates, is similar.

The Bass (1969) new product adoption model

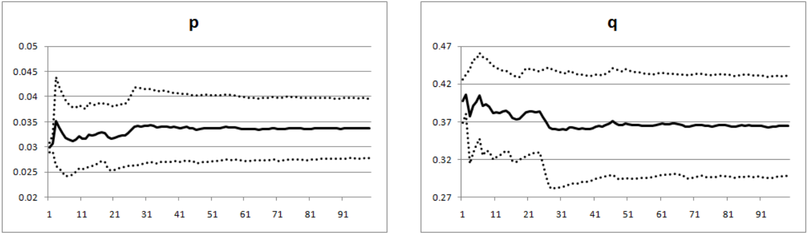
- Early example of epidemic models of innovation diffusion [Geroski, 2000]
- The speed of adoption at time t is a function of an external force (marketing effort) and an internal one (word-of-mouth):

$$h(t) = \frac{f(t)}{S(t)} = p + qF(t)$$

- Hugely influential in marketing, widely used in sales and technology adoption forecasting (estimation on early periods)
- The Bass model is an aggregate differential equation (cast in continuous time). Different translations into a representation for observed empirical and discrete data exist, with different assumptions about the error term to be added.
- These empirical formulations of the Bass model are generally estimated by NLLS. However, estimates are known to suffer from substantial bias and have non-standard asymptotic properties [Van den Bulte and Lilien, 1997; Boswijk and Franses, 2005].
- We give the Bass model a straightforward (but novel, to the best of our knowledge) interpretation in terms of individual hazard rates of a homogeneous population.
- In our Montecarlo experiments we use $p = 0.03$ and q=0.4. (The average value of p has been found to be 0.03, and is often less than 0.01. The average value of q has been found to be 0.38, with a typical range between 0.3 and 0.5 [Mahajan et al., 1995]).

Estimation of Bass model by MSM: Montecarlo results

Estimation sample: $t \in [1, 5]$

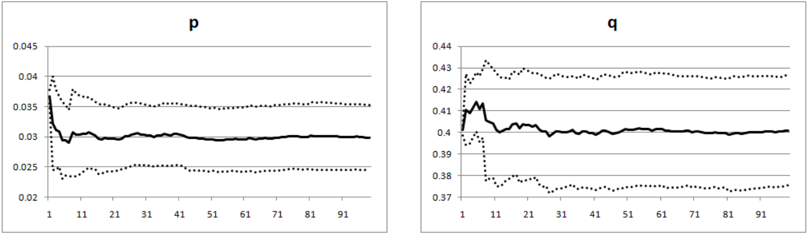


Cumulated mean over 100 replications ($\pm$ 1 std.dev.) $p = .03$; $q = .4$; 1,000 agents.

The estimator is biased (and inconsistent, as the bias does not converge to 0 as the population size increases –not shown here)

Estimation of Bass model by MSM: Montecarlo results

Estimation sample: all spells completed



Cumulated mean over 100 replications ($\pm$ 1 std.dev.) $p = .03$; $q = .4$; 1,000 agents.

The estimator is unbiased

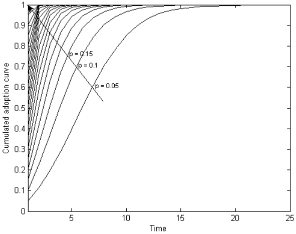
The ABBass model with local interaction

- We start from (micro version of) the Bass model with full connectivity, and replace the global interaction term $F(t)$ with a local interaction term $A(t)$:

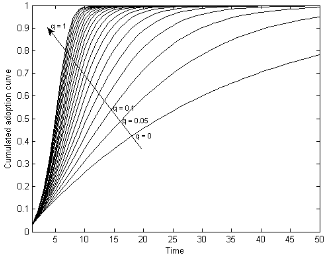
$$h(t) = p + qA(t)$$

- We consider a (bidirectional) random network with n average number of links
- Other AB Bass-type models with local interaction: [Alkemade and Castaldi, 2005] (one-dimensional grid), [Goldenberg et al., 2002] (bi-dimensional grid), [Garber et al., 2004; Delre et al.] (random networks),
- [Fibich and Gibori, 2010] derive exact aggregate diffusion dynamics for one-dimensional grids, where individuals are placed on a circle and can influence only their neighbors, and then extend their result to Cartesian grids of any dimensions: *they show that the spatial structure can have a large effect on the diffusion process*.

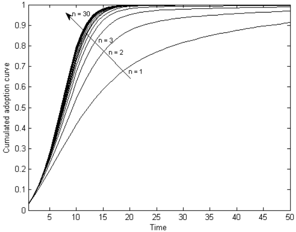
Sensitivity analysis



p



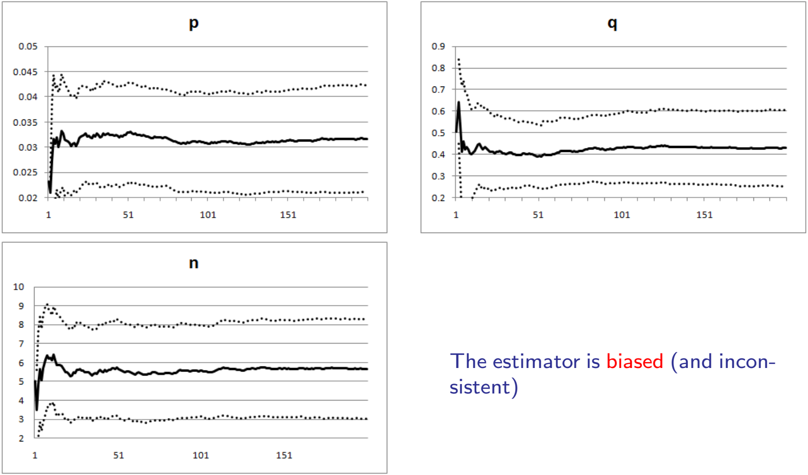
q



n

Estimation of ABBass model by MSM: Montecarlo results

Estimation sample: $t \in [1, 5]$

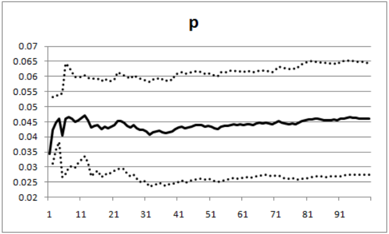


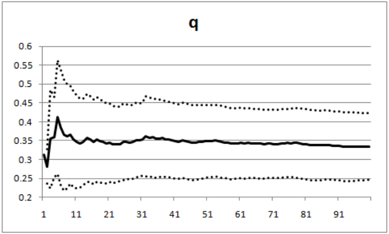
The estimator is biased (and inconsistent)

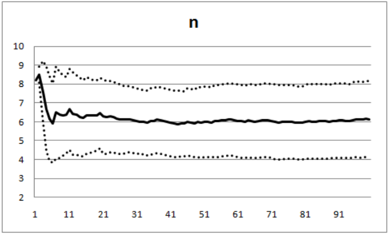
Cumulated mean over 200 replications ($\pm$ 1 std.dev.) $p = .03$; $q = .4$; $N = 5$; 1,000 agents.

Estimation of ABBass model by MSM: Montecarlo results

Estimation sample: all spells completed





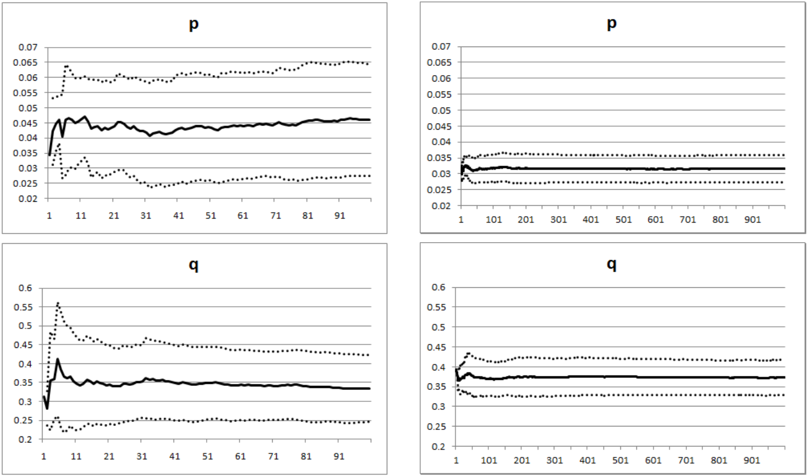


The estimator is biased (and inconsistent)

Cumulated mean over 100 replications ($\pm$ 1 std.dev.) $p = .03$; $q = .4$; $N = 5$; 1,000 agents.

Comparison of MSM and Path estimator of ABBass model

Estimation sample: all spells completed

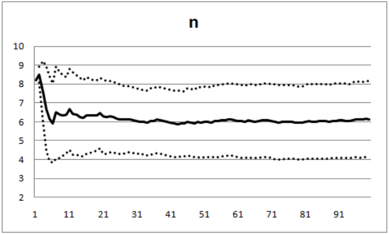


MSM Path

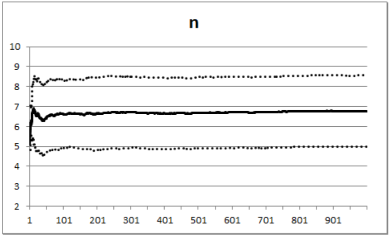
Cumulated mean ($\pm$ 1 std.dev.) $p = .03$; $q = .4$; $N = 5$; 1,000 agents.

Comparison of MSM and Path estimator of ABBass model

Estimation sample: all spells completed (cont'd)



MSM



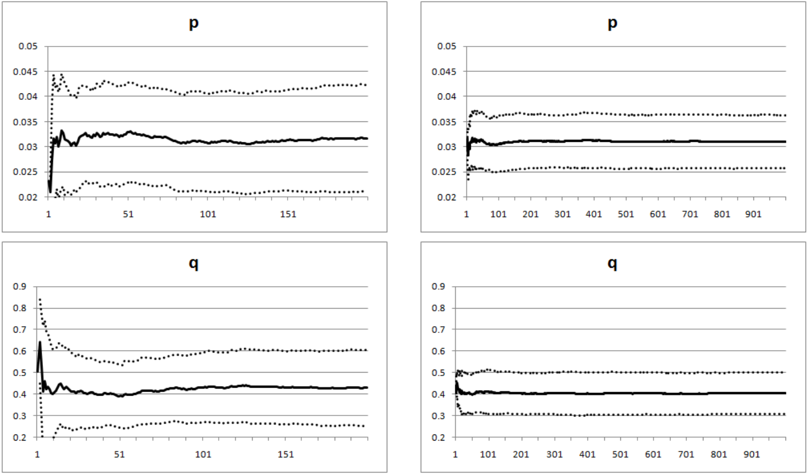
Path

Cumulated mean ($\pm$ 1 std.dev.) $p = .03$; $q = .4$; $N = 5$; 1,000 agents.

The path estimator is more precise than the MSM estimator.

Comparison of MSM and Path estimator of ABBass model

Estimation sample: $t \in [1, 5]$

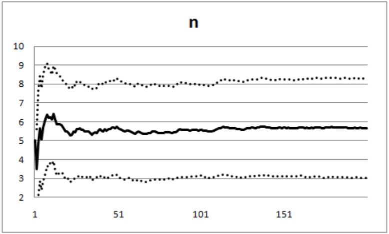


MSM Path

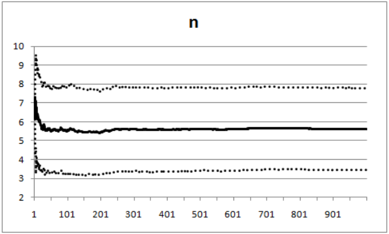
Cumulated mean ($\pm$ 1 std.dev.) $p = .03$; $q = .4$; $N = 5$; 1,000 agents.

Comparison of MSM and Path estimator of ABBass model

Estimation sample: $t \in [1, 5]$ (cont'd)



MSM



Path

Cumulated mean ($\pm$ 1 std.dev.) $p = .03$; $q = .4$; $N = 5$; 1,000 agents.

The path estimator is more precise than the MSM estimator.

Conclusions

- Estimation of ABMs is still rare.
- Estimates are often regarded with suspicion: the distinction between estimation and calibration, i.e. pure data tracking with no or little concerns for the properties of the estimators, being often weak.
- Proper estimation techniques exist. However, the distinctive features of ABMs make estimation difficult.
- AB specifications often entail many parameters: this increases the computational burden of numerical minimization algorithms and migh induce equifinality (under-identification). However, resorting to external information to pin down some of the parameters, as is common in calibration, does not help much (and might indeed even make things worse, as these parameters are estimated using different structural models).
- Interaction between the agents is likely to affect the properties of the estimators, especially when it has self-perpetuating, long-lasting effects (ergodicity of the process is questioned).
- The common assumption in AB modelling that real systems operate out of equilibrium further complicates estimation, as complete processes cannot be observed.
- Even when "everything can be observed", nonlinearities and asimmetries might induce permanent bias in the estimates.

Conclusions

Should we restrain from taking our models to the data, then?
NO!

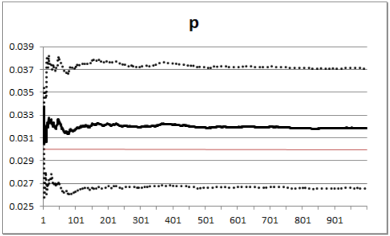
- People always try to estimate their model, irrespective of the difficulties (*sic!*)– e.g. the Bass model has been estimated hundreds of times over more than half a century, irrespective of the fact that it leads to biased and inconsistent estimates (and this has been known for at least the past 15 years).
- Even if the estimators loose their nice properties, the estimates might be close to the "true" values. Our estimates of an AB version of the Bass model are no more biased than traditional ones, and appear to discriminate well between values of the parameters with only a small effect on the output.

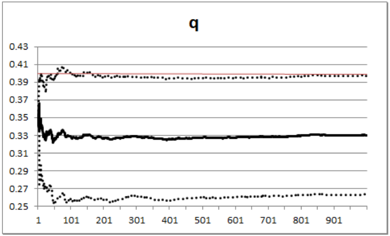
Neither we should abandon our models in favor of simpler ones only to take them to the data. The choice of the model specification should be guided primarily by considerations about the form of the *true* DGP: estimation of simpler but misspecified models leads to *wrong* interpretation of the coefficients.

Estimating Bass on ABBass

Montecarlo results. Estimation sample: $t \in [1, 5]$

- If the real DGP is ABBass...
- ...and we (incorrectly) estimate a Bass model...





Cumulated mean over 1,000 replications ($\pm$ 1 std.dev.) $p = .03$; $q = .4$; $N = 5$; 1,000 agents.

- ...we get a significantly lower estimate for the external influence parameter q
- (though the fit is still very high)

Thank you for your attention.

matteo.richiardi@unito.it