Online Appendix: Heterogeneous dynamics, aggregation and the persistence of economics shocks

Laura Mayoral

Institute for Economic Analysis, (CSIC)

November 1, 2011

Abstract

This Appendix contains materials not included in the main text. More specifically, it presents the proof of Theorem 1, reports the results of some Monte Carlo experiments that evaluate the finite sample properties of estimators based on aggregate and disaggregate data, and discusses some issues in connection with the aggregation of heterogeneous processes.

1. PROOF OF THEOREM 1

This section presents the proof of Theorem 1.

By definition of $IRF_{dis}\left(t,\ h\right)$,

$$IRF_{dis}\left(t,\ h\right) = E_I(IRF_{i,t}) = E_I\left(E\left(y_{it+h}|u_t = 1; z_{it-1}\right) - E\left(y_{it+h}|u_t = 0; z_{it-1}\right)\right)$$

and by Fubini's theorem (see Theorem 9.4.1. in Rosenthal, 2006),

$$= E\left(E_I\left(y_{it+h}\right)|u_t = 1; Z_{t-1}\right) - E\left(E_I(y_{it+h})|u_t = 0; Z_{t-1}\right)) \quad (1)$$

$$= E\left(Y_{t+h}|u_t = 1; Z_{t-1}\right) - E\left(Y_{t+h}|u_t = 0; Z_{t-1}\right)$$

$$= IRF_{AG}(t,h).\blacksquare \quad (2)$$

2. FINITE SAMPLE PROPERTIES OF IRF ESTIMATORS UNDER INDIVIDUAL HETEROGENEITY

This section explores whether it is possible to obtain good approximations to the population relationships established in Section 2 in the main text when the corresponding sample counterparts are employed. In order to do that, we have carried out two Monte Carlo experiments that estimate the impulse response to an aggregate shock using both aggregate and disaggregate data and estandard estimation techniques.

The first simulation explores whether the finite-sample properties of estimated IRFs deteriorate as the degree of heterogeneity of the underlying processes increases (for a given level of persistence). The second focuses on assessing how efficiency improves as the degree of desaggregation of the data increases.

In order to isolate the impact of an increase in heterogeneity from an increase in average persistence, different degrees of heterogeneity have been considered while keeping the level of persistence constant. To achieve this, the following approach has been adopted. The data has been generated according to the model:

$$y_{it} = a_i y_{it-1} + \rho_i u_t + \varepsilon_{it,}\ i = 1, ..., N,\ t = 1, ...T, \quad (3)$$

where $T = \{100, 400\}$, $N = \{100,\ 200\}$, u_t and ε_{it} are $i.N\left(0,1\right)$ and $i.N\left(0,\sigma_i^{\varepsilon}\right)$ random variables, respectively and σ_i^{ε} and ρ_i are draws from two independent uniform distributions in the interval $(0.5,\ 1.5)$. The autoregressive parameter a is distributed

as a $U(\omega_1, \omega_2)$, for different values of (ω_1, ω_2). These values determine both the degree of heterogeneity of the individual units and the level of persistence of the aggregate. In this exercise, the degree of heterogeneity will be determined by the standard deviation of a, σ_a, while the level of persistence will be measured by the cumulated impulse response up to lag 100, CIR(100).

To select (ω_1, ω_2), the following approach has been pursued. Firstly, three levels of persistence, denoted as LP I, II and III, have been considered. These persistence levels correspond to the case where $a_i = a$ for all i (i.e., there is no heterogeneity) and $a = \{0.85, 0.90, 0.95\}$, respectively. Persistence, as measured by the CIR(100), is equal to

$$LP \text{ I:} \qquad CIR(100) = \sum_{j=0}^{100} 0.85^j = 6.67.$$

$$LP \text{ II:} \qquad CIR(100) = \sum_{j=0}^{100} 0.9^j = 10.$$

$$LP \text{ III:} \qquad CIR(100) = \sum_{j=0}^{100} 0.95^j = 19.89.$$

Secondly, four values of σ_a have been considered, namely $\sigma_a^* = \{0, 0.01, 0.05, 0.1\}$. Next, the bounds of the uniform distribution (ω_1, ω_2) are chosen in such a way that the resulting σ_a reproduces the values of σ_a^* and the resulting set of heterogeneous processes imply a similar value of the CIR(100) as LP I to LP III. Under heterogeneity, $\text{CIR}(100) = \sum_{j=0}^{100} E(a^j) = \left(\frac{1}{\omega_2 - \omega_1}\right) \sum_{j=0}^{100} \left(\omega_2^{j+1} - \omega_1^{j+1}\right)/j + 1$. Then, the restriction on the standard deviation implies that $(\omega_2 - \omega_1) = \sqrt{12}\sigma_a$ while the condition on the CIR allows us to identify unique values for (ω_1, ω_2). TABLE AI reports these values for each of the levels of persistence and heterogeneity considered in this experiment.

TABLE AI. DISTRIBUTION OF a

$a \sim U(\omega_1, \omega_2)$									
		σ_a= 0.0		σ_a= 0.01		σ_a= 0.05		σ_a= 0.1	
	CIR(100)	ω_1	ω_2	ω_1	ω_2	ω_1	ω_2	ω_1	ω_2
LP I	6.67	0.85	0.85	0.83	0.866	0.7470	0.9202	0.6155	0.9619
LP II	10	0.90	0.90	0.882	0.9166	0.7900	0.9632	0.6455	0.9919
LP III	19.89	0.95	0.95	0.9300	0.9646	0.826	0.992	0.6700	1.00

Notes: This table reports the values of the bounds (ω_1, ω_2) of the uniform distribution that has been employed in Tables A2 and A3 to generate the values of a. These bounds depend on the Level of persistence (LP1 to LP III) as well as on the variance of a, σ_a^*=$\{0,0.01,0.05,0.1\}$.

The aggregate process, Y_t, has been computed as the simple average of the $y_{it}'s$

$$Y_t = \frac{\sum_{i=i}^{N} y_{it}}{N}. \tag{4}$$

As regards the estimation of Y_t, difficulties arise because it contains an infinite number of parameters. Berk (1974) and Lewis and Reinsel (1985) have shown that $\sqrt{T}$-consistent and asymptotically normally distributed estimates can be obtained by approximating the AR(∞) process by an AR(k) model, where k does not increase too quickly or too slowly.[1]

The lag length, k, is the key parameter in implementing procedures that approximate AR(∞) models in applications. Ng and Perron (1995) have argued that standard selection criteria (the AIC and the BIC) choose values of k, $\hat{k}$, that are proportional

[1] More specifically, k should verify an upper bound condition, $k^3/T \to 0$, and a lower bound one, $T^{1/2}\sum_{j=k+1}^{\infty}|A_j| \to 0$ as k, $T \to \infty$.

to $\log T$ and, so, do not verify the lower bound condition stated above.[2] Kuersteiner (2005) has shown that the general-to-specific (GTS) approach (Ng and Perron, 1995) can be used to produce a data-dependent selection rule such that the estimates obtained in the AR(k) model are consistent and asymptotically normal for the parameters of the underlying AR(∞) model. Then, the consistency and the asymptotic normality of aggregate IRF estimates in AR(∞) models follow from an application of the delta method.

AR(k) processes have been fitted to Y_t, where k was chosen according to the AIC and the GTS. Following Ng and Perron (1995), the maximum value of k, $k\max$, was set according to the rule $k\max(Y_T) \approx 10 * (T/100)^{0.25}$, which yields $kmax = \{10, 14\}$ for $T = \{100, 400\}$, respectively. A similar strategy has been followed to fit models for the individual $y_{it}s$. This time, $k\max$ was set to $k\max \approx (Y_T)/3$.

For each replication r, the estimated disaggregate and aggregate models have been employed to compute the corresponding IRFs, denoted as $\widehat{IRF_{dis,r}}$ and $\widehat{IRF_{AG,r}}$, respectively. $\widehat{IRF_{dis,r}}$ has been obtained as the mean of the individual IRFs while $\widehat{IRF_{AG,r}}$ is the sample analog of (11) in the main text.

Table AII reports the average mean squared error (MSE) of $\widehat{IRF_{dis}}$ and $\widehat{IRF_{AG}}$, which is defined as

$$\overline{MSE}_{dis} = R^{-1} \sum_{r=1}^{R} \left(h^{-1} \sum_{j=0}^{h} \left(\widehat{IRF_{dis,r}}(j) - IRF_{true,r}(j) \right)^2 \right), \tag{5}$$

and

$$\overline{MSE}_{AG} = R^{-1} \sum_{r=1}^{R} \left(h^{-1} \sum_{j=0}^{h} \left(\widehat{IRF_{AG,r}}(j) - IRF_{true,r}(j) \right)^2 \right), \tag{6}$$

where $R = 1000$ is the number of replications, $h = 20$ is the horizon of the IRF and

[2] In fact, bias terms arising as a consequence of the asymptotic misspecification of the model when these criteria are employed are of order $T^{-1/2}$, considerably more severe than the usual finite sample biases that are typically of order T^{-1}.

$IRF_{true,r}$ is the true IRF in replication r, given by

$$IRF_{true,r}(j) = N^{-1}\sum_{i=1}^{N} a_{i,r}^{j},\ j = 1, ..., h.$$

Only values obtained by selecting k using the GTS approach and $N = 100$ are reported. Those obtained by using the AIC and $N = 200$ were similar so they are omitted for the sake of brevity.

TABLE AII. MEAN SQUARED ERROR, GTS SELECTION METHOD

	σ_a= 0.0		σ_a= 0.01		σ_a= 0.05		σ_a= 0.1	
	$\overline{MSE}_{dis}$	$\overline{MSE}_{AG}$	$\overline{MSE}_{dis}$	$\overline{MSE}_{AG}$	$\overline{MSE}_{dis}$	$\overline{MSE}_{AG}$	$\overline{MSE}_{dis}$	$\overline{MSE}_{AG}$
				T=100				
LP I	0.063	0.234	0.064	0.259	0.061	0.239	0.056	0.210
LP II	0.168	0.461	0.166	0.470	0.153	0.456	0.101	0.328
LP III	0.569	1.091	0.570	1.138	0.303	0.748	0.138	0.432
				T=400				
LP I	0.013	0.061	0.014	0.061	0.0121	0.057	0.010	0.055
LP II	0.028	0.102	0.028	0.106	0.023	0.095	0.016	0.077
LP III	0.065	0.183	0.067	0.192	0.040	0.141	0.018	0.090

Notes: The disaggregate and aggregate processes have been generated according to (3) and (4), respectively. The number of disaggregate processes, N, is equal to 100 and the number of replications, R, is 1000. AR(k) processes have fitted to the data where k has been chosen according to the GTS approach. $\overline{MSE}_{dis}$ and $\overline{MSE}_{AG}$ are defined in (5) and (6), respectively.

Figures A1 to A3 present the average over the number of replications of IRF_{true}, $\widehat{IRF_{dis}}$ and $\widehat{IRF_{AG}}$ for LP I-III, respectively, for $T = 100$. Confidence bands at the 5% significance level are also reported. Four graphs are presented in each figure, corresponding to the different degrees of heterogeneity considered.[3] Graphs in the

[3]For each h, these bands are computed as the (0.250, 0.975) percentiles of the R replications of

upper left corner correspond to the case where there is no heterogeneity ($\sigma_a = 0$), while the remaining graphs display estimated IRFs under positive $\sigma_a s$.

Several interesting conclusions can be drawn from inspecting Figures A1 to A3 and Table AII. Firstly, the figures show that the average bias is in general small and fairly similar for the $\sigma_a = 0$ and the $\sigma_a > 0$ cases and for $\widehat{IRF_{dis}}$ and $\widehat{IRF_{AG}}$. Furthermore, confidence bands do not seem to vary much as σ_a increases. This suggests that the existence of individual heterogeneity does not hurt much the properties of the estimators. Confidence bands are considerably narrower for $\widehat{IRF_{dis}}$ than for $\widehat{IRF_{AG}}$, which implies that the use of disaggregate information can bring about important efficiency gains in the estimation of aggregate quantities. Similar conclusions are drawn by analyzing the average MSE: $\overline{MSE}_{dis}$ is always smaller than $\overline{MSE}_{AG}$, even in the absence of heterogeneity.

Secondly, increasing heterogeneity for a given level of persistence does not affect much $\overline{MSE}_{dis}$ unless persistence is high (LP III), in which case $\overline{MSE}_{dis}$ tends to decrease as heterogeneity increases. A similar pattern is observed for $\overline{MSE}_{AG}$: for moderate levels of persistence (LP I), increasing σ_a deteriorates $\overline{MSE}_{AG}$. However the opposite effect is found in more persistent situations. For instance, for LP III with $T = 100$, the $\overline{MSE}_{AG}$ for $\sigma_a = 0.1$ is less than half the $\overline{MSE}_{AG}$ obtained for the case of no heterogeneity ($\sigma_a = 0$). This evidence is surprising and at odds with the common belief that individual heterogeneity hurts the properties of aggregate estimators.

Thirdly, increasing the level of persistence always increases the MSE. This increase may be due to the well-known downward bias affecting the OLS estimator of AR coefficients when persistence is high. Interestingly, the increase tends to be smaller in the presence of heterogeneity. For instance, for $\sigma_a = 0$ and $T = 100$, the MSE of $\widehat{IRF_{dis}}$ $\left(\widehat{IRF_{AG}}\right)$ increases by a factor of 9 (4.5) when moving from LP I to LP

the IRF(h) .

III. However, for $\sigma_a = 0.1$, the MSE of $\widehat{IRF_{dis}}$ $\left(\widehat{IRF_{AG}}\right)$ rises only by a factor of 2.5 (2) from LP I to LP III. Finally, both the bias and the MSE drop considerably when larger sample sizes are considered.

(Figure A1 about here)

(Figure A2 about here)

(Figure A3 about here)

An interesting conclusion of the previous exercise is that important efficiency gains are obtained when disaggregate data is employed to estimate the aggregate IRF. However in applications, highly disaggregate data is sometimes not available. We have carried out a second experiment to assess how efficiency improves as more disaggregate data is used. To do that, we have considered a similar DGP as before where N, the number of individual processes, is set equal to 200. Next, this data has been aggregated in several steps: Level of aggregation 5 (LAgg 5) contains 100 processes, each computed as the simple average of two of the original processes. LAgg 4 to 1 have been computed in an analogous way, such that they contain 50, 10, 5 and 1 process, respectively. Thus, LAgg 1 is simply the aggregate process defined in (4). Table AIII presents the MSE obtained by estimating these data following a similar strategy as in the previous exercise. We focus on LP III and consider the same levels of heterogeneity as before, namely, $\sigma_a = \{0,\ 0.01,\ 0.05,\ 0.1\}$.

Not surprisingly, the MSE decreases considerably as more disaggregate data is employed. Comparing the MSE of LAgg 1 with that of LAgg 5 for $T = 100$, it is seen that the latter is around 40 to 60% smaller. Interestingly, even when highly aggregate data is considered, as in LAgg 2, where each of the series is the result of aggregating 40 AR(1) heterogeneous processes, efficiency gains of around 20% are obtained with respect to LAgg 1. This fact stresses the convenience of using disaggregate data

whenever is available.

TABLE AIII. MSE AT DIFFERENT AGGREGATION LEVELS

	σ_a= 0.0	σ_a= 0.01	σ_a= 0.05	σ_a= 0.1
			T=100	
LAgg 1 (1 process)	1.144	1.071	0.774	0.436
LAgg 2 (5 processes)	0.946	0.948	0.631	0.353
LAgg 3 (10 processes)	0.919	0.883	0.605	0.333
LAgg 4 (50 processes)	0.769	0.738	0.462	0.226
LAgg 5 (100 processes)	0.666	0.639	0.374	0.168
			T=400	
LAgg 1 (1 process)	0.213	0.173	0.146	0.094
LAgg 2 (5 processes)	0.188	0.159	0.133	0.091
LAgg 3 (10 processes)	0.180	0.152	0.122	0.079
LAgg 4 (50 processes)	0.133	0.111	0.072	0.037
LAgg 5 (100 processes)	0.101	0.084	0.047	0.022

Notes: The disaggregate data has been generated according to (3), LP III and different levels of sigma. The number of disaggregate processses, N, is equal to 200. LAgg 1 to 5 aggregate the original data in 1, 5, 10, 50 and 100 series, respectively. The number of replications, R, is 1000. AR(k) processes have fitted to the data where k has been chosen according to the GTS approach. The MSE is computed according to (5).

3. AGGREGATION AND THE LAW OF LARGE NUMBERS

Throughout the main text it is assumed that the aggregate model is defined as the expected value of the individual processes. However, in applications aggregate data, denoted as $\bar{Y}_{Nt}$ henceforth, is constructed as a (weighted) average of the individual data. Hence, $\bar{Y}_{Nt}$ would approximately follow the aggregate model in (8) in the main

text if a LLN relating Y_t and $\bar{Y}_{Nt}$ holds. However, in some cases such a LLN might not hold (Forni and Lippi, 1997, p. 17). Since the applicability of the results obtained in Section 2 relies on this convergence, it is worth considering this issue in more detail.

For simplicity, we assume that the aggregate data $\bar{Y}_{Nt}$ is constructed as a simple average of a large number of individual processes

$$\bar{Y}_{Nt} = \frac{\sum_{i=1}^{N} y_{it}}{N},$$

where y_{it} is defined as in (1) in the main text, with $b_i = 0$ for all i. $\bar{Y}_{Nt}$ can be written as the sum of two terms

$$\bar{Y}_{Nt} = \frac{1}{N}\left(\sum_{i=1}^{N} \frac{\varepsilon_{it}}{(1 - a_i L)} + \sum_{i=1}^{N} \frac{\rho_i u_t}{(1 - a_i L)}\right), \tag{7}$$

that will be referred to as the idiosyncratic and common components, respectively. Then, $\bar{Y}_{Nt}$ would be a good approximation of the aggregate model derived in Section 3 provided that a LLN applies, such that $\bar{Y}_{Nt}$ and $Y_t = E_I(y_t)$ are close for large N.

The holding of a LLN relating $\bar{Y}_{Nt}$ and Y_t hinges on whether the limit of $\bar{Y}_{Nt}$ when $N \to \infty$ is stationary or not. So, before considering the convergence of $\bar{Y}_{Nt}$ and Y_t, the asymptotic properties of $\bar{Y}_{Nt}$ as N increases should be reviewed. This issue has been analyzed by Zaffaroni (2004) and we briefly summarize the results that are relevant to the problem considered here. The asymptotic behavior of $\bar{Y}_{Nt}$ critically depends on the properties of the distribution of a around 1. As shown by Granger (1980), if the support of a is given by $[\omega_1, \omega_2]$ with $\omega_2 < 1$, the corresponding aggregate process is I(0), for any shape of the distribution of a. On the other hand, if $\omega_2 = 1$ and the distribution of a is such that $P(a = 1) > 0$, then $\bar{Y}_{Nt}$ converges to an I(1) random variable. An interesting intermediate case arises whenever $\omega_2 = 1$ and a belongs to a family of absolutely continuous distributions such that $P(a = 1) = 0$. To characterize the convergence in this case, Zaffaroni (2004) considers the following semiparametric

specification of the density of $a \in (0,1)$ around unity,[4]

$$f(a) \sim c_b (1-a)^{-b}, \text{ as } a \to 1, \ 0 < c_b < \infty, \ b \in [0,1)$$

In this case, $\bar{Y}_{Nt}$ converges to a stationary random variable provided $b < 0.5$ and to a nonstationary one otherwise. Interestingly, if $0 \leq b < 0.5$, the limit of $\bar{Y}_{Nt}$ is a long-memory process and if $b > 0$, the limit process can be characterized as a fractionally integrated process with order of integration $d = b$.

Under similar assumptions as the ones adopted in this paper, Zaffaroni (2004) shows that provided the limit of $\bar{Y}_{Nt}$ is stationary, a strong LLN holds and $\bar{Y}_{Nt} \xrightarrow{L_2} E_t(y_t) = Y_t$. In this case, the idiosyncratic component converges almost surely to zero while the common component converges in L_2 to the corresponding expectation[5]

$$N^{-1} \sum_{i=1}^{N} \frac{\rho_i u_t}{(1-a_i L)} \xrightarrow{L_2} E_I \left(\frac{\rho u_t}{1-aL} \right) = \sum_{j=0}^{\infty} E_I \left(a^j \right) u_{t-j}.$$

However, whenever the limit of $\bar{Y}_{Nt}$ is a nonstationary random variable, the convergence above fails: the idiosyncratic component no longer vanishes because the variance of $N^{-1} \sum_{i=1}^{N} \frac{\varepsilon_{it}}{(1-a_i L)}$ tends to infinity. On the other hand, the common component does not converge to its expected value because neither the Bochner nor the Pettis integral of this component exist.

In principle, this could be a major drawback for the results established in Section 2. If nonstationary variables are observed, $\bar{Y}_{Nt}$ might not be a good proxy for Y_t. Then, one should not expect the persistence estimates obtained with aggregate data, $\bar{Y}_{Nt}$,

[4] This condition is semiparametric because the behavior of the density for any given interval $[0, \gamma_2]$ with $\gamma_2 < 1$ is unspecified. Standard distributions, such as the Uniform or the Beta, are contained in this specification by setting $b = 0$ and $b \geq 0$, respectively.

[5] The expectation of the idiosyncratic component is taken with respect to the Pettis integral (see Uhlig, 1986). This is because the Bochner integral (which extends the definition of the Lebesgue integral to functions taking values in a Banach space) of that component may not exist. This is the well known *measurability problem* (Judd, 1985).

to be close to those obtained with the corresponding disaggregate variables. Notice, however, that this problem has an easy solution. Taking first differences from the original aggregate data, $\bar{Y}_{Nt}$, we obtain

$$(1-L)\,\bar{Y}_{Nt} = \frac{1}{N}\left(\sum_{i=1}^{N}\frac{(1-L)\,\varepsilon_{it}}{(1-a_iL)} + \sum_{i=1}^{N}\frac{(1-L)\,\rho_i u_t}{(1-a_iL)}\right), \tag{8}$$

and, in this case, the same results as in the case where the limit of $\bar{Y}_{Nt}$ is stationary are recovered, that is, the idiosyncratic component in (8) converges to zero while the common one converges to the corresponding expectation. Thus, it holds that

$$(1-L)\,\bar{Y}_{Nt} \xrightarrow{L_2} (1-L)\,Y_t, \tag{9}$$

where $(1-L)\,Y_t$ is the first difference of $Y_t = E_I(y_t)$ and is a stationary process. Thus, whenever nonstationarity is detected, the usual procedure of first differentiating the data would be sufficient in order to guarantee the convergence of $(1-L)\,\bar{Y}_{Nt}$ to $(1-L)\,Y_t$. The IRF of Y_t can be estimated by first estimating the IRF associated with $(1-L)\,\bar{Y}_{Nt}$, and, then, cumulating the corresponding values. That is,

$$\widehat{IRF_{AG}}(t,h) = \sum_{j=1}^{h}\widehat{IRF_{(1-L)\bar{Y}_N}}(j,t),$$

where $\widehat{IRF_{AG}}(t,h)$ and $\widehat{IRF_{(1-L)\bar{Y}}}(j,t)$ denote the estimates of the IRFs associated with Y_t and with $(1-L)\,Y_t$, respectively.

REFERENCES

Berk, K.N. (1974), "Consistent autoregressive spectral estimates," *The Annals of Statistics* 2, pp. 489-502.

Forni M. and M. Lippi (1997), Aggregation and the microfoundations of dynamic macroeconomics, Oxford University Press Inc., New York.

Granger, C.W.J. (1980), "Long memory relationships and the aggregation of dynamic models", *Journal of Econometrics* 14, pp. 227-238.

Judd, K. L. (1985), "The law of large number with a continuum of IID random variables," *Journal of Economic Theory* 35, 19-25.

Lewis, R. and G. Reinsel (1985), "Prediction of multivariate time series by autoregressive model fitting," *Journal of Multivariate Analysis* 64, pp. 393-411.

Ng, S. and P. Perron, (1995), "Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag," *Journal of the American Statistical Association* 90, pp. 268-281.

Rosenthal, J.S (2006), *A First Look at Rigorous Probability Theory.* World Scientific Publishing Co.

Uhlig, H. (1996), "A law of large numbers for large economies," *Economic Theory* 8, 41-50.

Zaffaroni, P. (2004), "Contemporaneous aggregation of linear dynamic models in large economies," *Journal of Econometrics* 120, pp. 75-102.

FIGURES

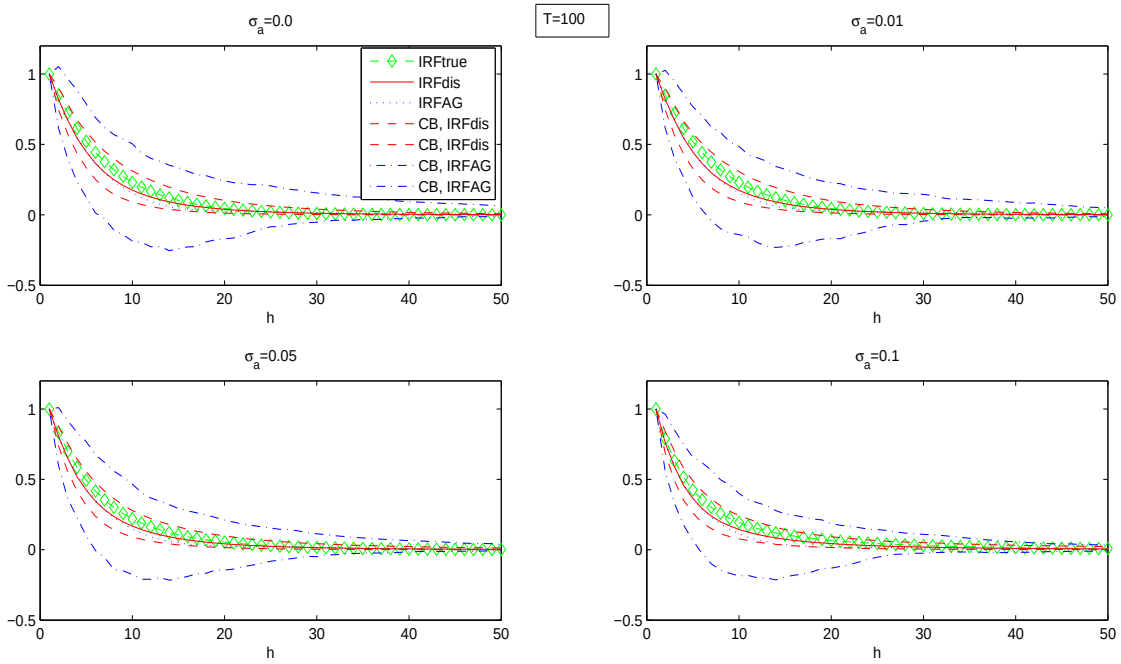

FIG A1. Estimated IRFs, LP I. T=100.

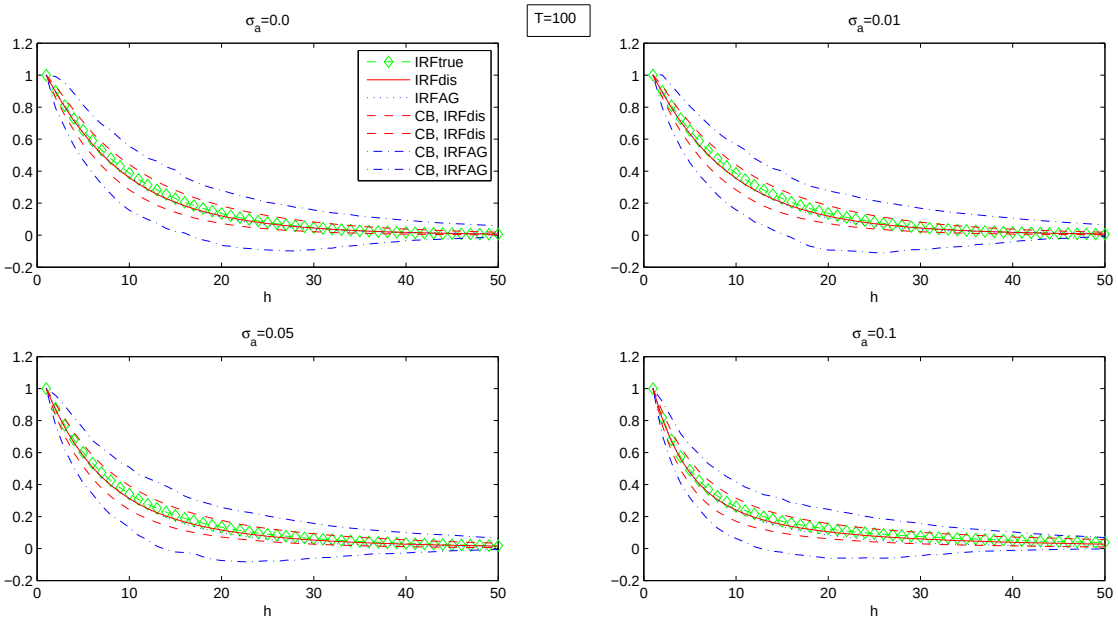

FIG A2. Estimated IRFs, LP II. T=100.

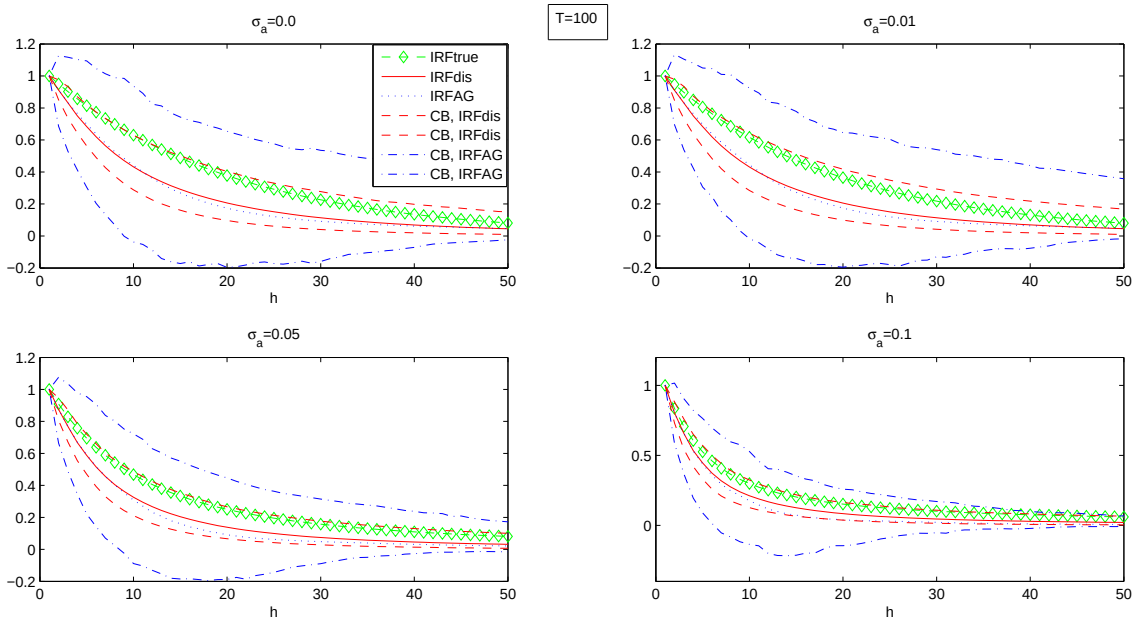

FIG A3. Estimated IRFs, LP III T=100.