# Stability Analysis of Heterogeneous Learning in Self-Referential Linear Stochastic Models

Chryssi Giannitsarou*†
London Business School

Preliminary, May 2001

**Abstract**

There is by now a large literature characterizing conditions under which learning schemes converge to rational expectations equilibria (REEs). A number of authors have claimed that these results are dependent on the assumption of homogeneous agents and homogeneous learning. We study stability analysis of REEs under heterogeneous adaptive learning, for the broad class of self-referential linear stochastic models. We introduce three types of heterogeneity related to the way agents learn: different perceptions, different degrees of inertia in updating, and different learning algorithms. We provide general conditions for local stability of an REE. Even though in general hetereogeneity may lead to different stability conditions, we provide applications to various economic models where the stability conditions are identical to the conditions required under aggregation. This suggests that heterogeneity may affect the stability of the learning scheme but that in most models aggregation works locally.

*JEL classification:* D83, C62
*Key words:* heterogeneity, learning, local stability

*I am grateful to Albert Marcet for valuable discussions, and to Andrew Scott and Flavio Toxvaerd for their comments and suggestions. Part of this paper was completed during a visit to the Economics Department of Universitat Pompeu Fabra. I thank them for their hospitality. All errors are mine.

†*Address for correspondence:* London Business School, Economics Department, Regent's Park, London NW1 4SA, United Kingdom. *E-mail*: CGiannitsarou@london.edu.

## 1. Introduction

A significant part of the rapidly developing learning literature has concentrated on characterizing conditions under which learning schemes converge to rational expectations equilibria (REEs). The importance of these contributions has various dimensions. Not only does learning provide a conceptual improvement from the by now standard assumption of rational expectations (RE), it also serves as a test of robustness of equilibria to expectational errors, or as a selection mechanism in models with multiple equilibria, and it has made it possible to explain economic phenomena that could not be tackled using RE methodology. These advantages have been stressed by numerous authors over the last fifteen years. Nevertheless, several points of the learning approach have been criticised. Perhaps the most important one, is the assumption of the representative agent.

In macroeconomic theory, the assumption of the representative agent has often been criticised[1], not only because it is unrealistic, but also because it might yield misleading conclusions regarding the dynamics and behaviour of an economy. Furthermore, in learning models, apart from the structural heterogeneity that may arise within the economy, there is the additional issue of the degree of expectational coordination among the agents. Although the importance of this point has been stressed, it has been somewhat ignored, perhaps because of the early indications in the literature that have been supportive of the representative agent, and also due to the technical simplicity of analysing the stability under this assumption. However, the small number of contributions concerning heterogeneous learning, especially the more recent ones, give no clear indications but, on the contrary, a certain amount of ambiguity. Some authors have shown that heterogeneity does not matter (Bray & Savin (1986), Sargent (1993), Evans & Honkapohja (1996)) while others show that it does matter (Marcet & Sargent (1989b), Barucci (1997), Franke & Nesemann (1999) and Evans, Honkapohja & Marimon (2000)). The source of ambiguity regarding the plausibility of the representative agent is the lack of a general systematic study of heterogeneous learning. With the exception of Marcet & Sargent (1989b), the stability results obtained in the above papers are very much dependent either on the structural specifics of the models, or on the particular and not always well justified learning algorithm that is employed.

In this paper, I present an analysis of the local asymptotic properties of heterogeneous learning for the broad class of self-referential linear stochastic models. The term *heterogeneous learning* is used to emphasise that it refers to differences in the ways agents learn, and not structural heterogeneities of the model. The

[1] For an enlightening critisism on the representative agent assumption see Hahn & Solow (1997).

purpose of this choice is to explore exactly what would happen when the single asymmetry of the agents is how they learn, as structural heterogeneity would involve unnecessary complications, that could remove the focus from the comparison with the representative 'learner'. I study three types of heterogeneity: agents that (i) have different expectations (or perceptions) (ii) have different degrees of inertia in updating and (iii) use different learning rules. The analysis consists of deriving conditions for local asymptotic stability of rational expectations equilibria (REEs) under the heterogeneous algorithm and comparison of these with the stability conditions for the learning rule of the representative agent.

Interestingly, it turns out that for the case of heterogeneous expectations, when the agents use the recursive least squares learning scheme, the conditions for local convergence of heterogeneous and homogeneous learning are always identical. However, the stability conditions for the remaining types of heterogeneity are not necessarily the same as the ones under homogeneous learning, for the general setup. For this reason, the results are applied to four sub-classes of the class of self-referential linear stochastic models. These cover a wide range of standard macroeconomic models. For these sub-classes it can be shown that the conditions for all the types of heterogeneity are identical with the ones of the homogeneous case.

The paper consists of the following sections. First I describe the general formulation of the model and the main tools for analysing stability of learning models. Second I briefly discuss the convergence and the stability properties under homogeneous learning, and in particular for the recursive least squares and stochastic gradient schemes that have been the most popular learning rules used in the literature. Next I proceed with the stability analysis of heterogeneous learning for the three types of heterogeneity, and last I apply the stability results to four reduced form examples. Closing comments follow.

## 2. The general setup

For completeness, I first give the general description of the class of models to be studied, i.e. self-referential linear stochastic models (SRLS models). Following the notation of Marcet and Sargent (1989a), the model at time $t$ is described by an $n-$dimensional vector of random variables $z_t \in \mathbb{R}^n$. Suppose that $z_{1t} \in \mathbb{R}^{n_1}$ is the subvector of $z_t$ which contains the variables that the agents are interested in predicting, and that $z_{2t} \in \mathbb{R}^{n_2}$ is the vector of variables that are relevant for predicting $z_{1t}$. The agents believe in the following perceived law of motion of the variables

$$z_{1t} = \Phi_t' z_{2t-1} + \eta_t$$

where $\eta_t$ is a vector of white noise errors, orthogonal to all past $z_2$'s, and with zero mean. $\Phi_t$ is an $n_2 \times n_1$ matrix of parameters. The actual law of motion for $z_t$ is then

$$z_t = \begin{pmatrix} z_{1t} \\ z_{1t}^c \end{pmatrix} = \begin{pmatrix} \mathbf{0} \;\; T(\Phi_t) \\ A(\Phi_t) \end{pmatrix} \begin{pmatrix} z_{2t}^c \\ z_{2t} \end{pmatrix} + \begin{pmatrix} V(\Phi_t) \\ B(\Phi_t) \end{pmatrix} \cdot \mathbf{u}_t$$

where the superscript $c$ denotes the complement of the relevant vector. The rational expectations equilibria (REEs) of the SRLS model belong to the set of the fixed points of the $T-$ map, i.e. solutions of the equation $T(\Phi_f) = \Phi_f$. Therefore, the study focuses on analysing the asymptotic local properties of such solutions, $\Phi_f$.

This setup covers a wide range of macroeconomic models. In particular, any linear model that can be written in a reduced form that contains lags of the endogenous variables, lags of exogenous variables, and lagged or future expectations of future values of the endogenous variables, can be studied within this framework. For example, consider the general reduced form

$$y_t = \mu + \sum_{i=1}^{l} \alpha_i y_{t-i} + \sum_{j=1}^{m} \sum_{k=1}^{n} \beta_{jk} E_{t-j}^* y_{t-j+k} + \sum_{s=1}^{r} \gamma_s w_{s,t} \tag{2.1}$$

where $y_t$ is a vector of endogenous variables, $E_{t-j}^* y_{t-j+k}$ is the expectation of $y_{t-j+k}$ formed by the agents at time $t-j$, and $w_{s,t} = \rho_s w_{s,t-1} + \varepsilon_{s,t}$ are vectors of exogenous variables. The exact specification of the vectors $z_{it}$ depends on the model at hand. Several examples can be found in Marcet & Sargent (1989a), and Evans & Honkapohja (2001). Furthermore, four special cases of (2.1) will be studied in section 5 to illustrate how the stability results obtained here can be applied.

Suppose now that agents' beliefs $\Phi_t$ are updated according to the following adaptive learning algorithm:

$$\theta_t = \theta_{t-1} + \alpha_t Q(\theta_{t-1}, z_{2t-1})$$

where $\theta_t$ is a vector containing the (vectorised) beliefs of the agents $\Phi_t$ and possibly other auxiliary parameters that are used for updating, and $\alpha_t$ and $Q(\cdot,\cdot)$ satisfy some technical assumptions[2]. If the necessary assumptions are satisfied, the learning algorithm can be associated with the ordinary differential equation (henceforth ode)

$$\frac{d\theta}{d\tau} = h(\theta)$$

[2]For completeness, these assumptions are stated in appendix A.

where $h(\theta) = \lim_{t\to\infty} E\left[Q(\theta, z_{2t}(\theta))\right]$. The following results have been established in stochastic approximation theory: (a) If this ode has an equilibrium point $\theta^*$ which is locally asymptotically stable, then the algorithm converges to $\theta^*$ with some probability which is bounded from below by a sequence of numbers tending to one (Evans & Honkapohja, 1998a), (b) If $\theta^*$ is not an equilibrium point, or if it is not a locally asymptotically stable equilibrium point of the ode, then the algorithm converges to $\theta^*$ with probability zero (Ljung, 1977).

If the ode method can be applied, then the convergence and the local asymptotic stability of an equilibrium $\theta^*$ of the learning algorithm are determined by the local asymptotic stability of the associated ode, which in turn is determined by the stability of the matrix $J(\theta^*) = \left.\frac{\partial \text{ vec}h(\theta)}{\partial \text{ vec}\theta}\right|_{\theta=\theta^*}$. Therefore the conditions required for convergence and stability of the learning algorithm (henceforth stability conditions) are derived by imposing that $J(\theta^*)$ is a stable matrix[3].

## 3. Homogeneous learning

The bulk of the adaptive learning literature deals with stability analysis and provides results under the assumption that agents are homogeneous in the way the learn the relevant parameters of the economy. Typically, the agents are assumed to have some basic knowledge of econometrics, such that the parameters $\Phi_t$ can be interpreted as ordinary least squares estimates based on data up to time $t-1$. Recursive least squares has been extensively used, mainly for two reasons. First, because it is a reasonable and statistically efficient learning rule. Second, because, as Marcet & Sargent (1989a) show, the technical difficulty when studying the convergence of the algorithm can be reduced considerably.

A popular alternative learning rule is the stochastic gradient algorithm[4] (see Sargent (1993), Kuan & White (1994), Barucci & Landi (1997), Evans & Honkapohja (1998b) and Heinemann (2000)). The essential difference between stochastic gradient learning and recursive least squares learning is that the former is a gradient type algorithm, while the latter is a Newton type algorithm (i.e. it uses information on second moments). Naturally, stochastic gradient learning is computationally less complex than recursive least squares learning, and could therefore be considered a more plausible learning device for economic agents from a behavioural point of view, as all the above authors point out. I now turn to a brief description of the two algorithms of interest.

[3] A matrix is called stable if all its eigenvalues have negative real parts.

[4] Barucci and Landi (1997) refer to it as 'least mean squares learning'.

**Recursive least squares learning.** Using the notation of section 2, the recursive least squares learning algorithm is given by

$$\begin{aligned}\Phi_t &= \Phi_{t-1} + \alpha_t R_{t-1}^{-1} z_{2t-1}\left[z'_{2t-1}\left(T(\Phi_{t-1})' - \Phi'_{t-1}\right) + u'_{t-1}V(\Phi_{t-1})'\right] \\ R_t &= R_{t-1} + \alpha_t\left[z_{2t-1}z'_{2t-1} - R_{t-1}/t\alpha_t\right]\end{aligned}$$

Marcet and Sargent (1989a) show that the associated ode is the vectorised version of the following ode[5]

$$\begin{aligned}\frac{d\Phi}{d\tau} &= R^{-1}M(\Phi)\left[T(\Phi)' - \Phi\right] \\ \frac{dR}{d\tau} &= M(\Phi) - R\end{aligned}$$

where $M(\Phi) = \lim_{t\to\infty} Ez_{2t}(\Phi)z_{2t}(\Phi)'$. The local stability of an REE is entirely determined by the local stability of the ode at the REE

$$\frac{d\text{vec}\Phi}{d\tau} = \text{vec}\left(T(\Phi)' - \Phi\right)$$

The Jacobian of the rhs of the ode is

$$J^{LS}(\Phi) = \frac{d\ \text{vec}\left(T(\Phi)' - \Phi\right)}{d\ \text{vec}\Phi} = L(\Phi) - I_{n_1 n_2}$$

where $L(\Phi) = d\ \text{vec}(T(\Phi)')\,/d\ \text{vec}\Phi$. The local asymptotic stability of an REE $\Phi_f$ under least squares learning is determined by the stability of the matrix $J^{LS}(\Phi_f)$ : the least squares algorithm converges to the locally asymptotically stable REE if and only if the real parts of the eigenvalues of $J^{LS}(\Phi_f)$ are strictly negative (Marcet & Sargent, 1989a).

**Stochastic gradient learning.** The stochastic gradient algorithm is given by

$$\Phi_t = \Phi_{t-1} + \alpha_t z_{2t-1}\left[z'_{2t-1}\left(T(\Phi_{t-1})' - \Phi_{t-1}\right) + u'_{t-1}V(\Phi_{t-1})'\right]$$

Barucci & Landi (1997) show that the associated ode is

$$\frac{d\text{vec}\Phi}{d\tau} = \text{vec}\left[M(\Phi)\left(T(\Phi)' - \Phi\right)\right]$$

[5] To convert this algorithm to the standard general form described in the previous section, one has to perform the timing transform $S_t = R_{t+1}$. This change does not alter the asymptotic behaviour of the algorithm, and therefore, although technically more precise, will be avoided here for consistency with the existing literature.

The Jacobian of rhs of the ode is[6]

$$
\begin{aligned}
J^{SG}(\Phi) &= \frac{d\ \mathrm{vec}\left[M(\Phi)\cdot(T(\Phi)'-\Phi)\right]}{d\ \mathrm{vec}\Phi} \\
&= \left[(T(\Phi)'-\Phi)'\otimes I\right]\cdot\frac{d\mathrm{vec}M(\Phi)}{d\ \mathrm{vec}\Phi}+\left[I\otimes M(\Phi)\right]\cdot J^{LS}(\Phi)
\end{aligned}
$$

The local asymptotic stability of an REE $\Phi_f$ under stochastic gradient learning is determined by the stability of the matrix $J^{SG}(\Phi_f)=[I\otimes M(\Phi_f)]\cdot J^{LS}(\Phi_f)$ : the stochastic gradient algorithm converges to the locally asymptotically stable REE if and only if the real parts of the eigenvalues of $J^{SG}(\Phi_f)$ are strictly negative (Barucci & Landi, 1997).

## 4. Heterogeneous learning

Unfortunately, homogeneous learning, whether it is with least squares, stochastic gradient or any other algorithm, suffers from at least the same problems as the representative agent in macroeconomic theory in general. In particular for learning, behind the representative agent lies the assumption that either (a) everybody coordinates with each other to act (learn) in precisely the same way or that (b) although the agents might learn in different ways, it suffices to study the actions of the agents on average. The first case is arguably unrealistic unless some cooperative element is introduced[7], while the second should not be trusted unless it can be *shown* rigorously that analysing the heterogeneous case is indeed equivalent to studying the learning of the average agent. The present work deals with examining the validity of assumption (b).

This section consists of a description and analysis of convergence of three types of heterogeneity that may arise as a natural consequence of the agents' limited rationality in models of learning. In particular, the heterogeneity studied here is related to the way agents *learn,* rather than to the structure of the model. It is assumed that the economy consists of a continuum of agents of measure one, and there are two types of agents, type A and type B, of measure $\psi$ and $1-\psi$ respectively. In contrast to the homogeneous case, here type A and B agents form expectations according to

$$
\begin{aligned}
z_{1t} &= \Phi'_{At}z_{2t-1}+\eta_t \\
z_{1t} &= \Phi'_{Bt}z_{2t-1}+\eta_t
\end{aligned}
$$

[6] For the derivation see appendix B.

[7] Evans & Guesnerie (1999) show that it is possible to trigger complete coordination of expectations on some perfect foresight path when there is common knowledge among the agents that the solution is near the path.

respectively, which implies that $E^*_{At}(z_{1t}) = \Phi'_{At} z_{2t-1}$ and $E^*_{Bt}(z_{1t}) = \Phi'_{Bt} z_{2t-1}$. Then

$$E^*_t(z_{1t}) = \left[\psi \Phi'_{At} + (1-\psi)\Phi'_{Bt}\right] z_{2t-1}$$

Let $\mathbf{\Phi}_t = (\Phi_{At}, \Phi_{Bt})$ be an $n_2 \times 2n_1$ matrix containing the estimates of the parameters for both agents at time $t$ and $g(\mathbf{\Phi}_t) = \psi\Phi_{At} + (1-\psi)\Phi_{Bt}$ be the function representing the weighted average of the parameter estimates of the two agents. Then the (average) perceived law of motion is

$$z_{1t} = g(\mathbf{\Phi}_t)' \, z_{2t-1} + \eta_t$$

and the true law of motion is given by

$$z_t = \begin{pmatrix} z_{1t} \\ z^c_{1t} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & T(g(\mathbf{\Phi}_t)) \\ & A(g(\mathbf{\Phi}_t)) \end{pmatrix} \begin{pmatrix} z^c_{2t} \\ z_{2t} \end{pmatrix} + \begin{pmatrix} V(g(\mathbf{\Phi}_t)) \\ B(g(\mathbf{\Phi}_t)) \end{pmatrix} \cdot \mathbf{u}_t$$

Note that the mapping $T$ is actually not altered; what changes compared to the homogeneous case is the argument at which it is evaluated. Clearly the REEs are not altered either, since under RE $E_t(z_{1t}) = E_{At}(z_{1t}) = E_{Bt}(z_{1t}) = \Phi_f z_{2t-1}$.

Before proceeding with the description of the types of heterogeneity to be analysed, I will briefly discuss two types of heterogeneity which do not fit into the above framework. First is the case of agents with asymmetric or private information, i.e. a case where the groups of agents have access only to subsets of the relevant state variables. Second is the case where some part of the population persistently misspecifies the model, by always ignoring some variables that actually influence the endogenous state variables[8]. It is beyond the scope of the current work to give a thorough discussion of the conceptual implications of these two assumptions. However, it should be mentioned that there are models that fit these descriptions, as discussed in Marcet & Sargent (1989b) for the case of private information, an in Evans & Honkapohja (2001, chapter 13) for the case of misspecifications. Formally, both these cases can be described and analysed within the framework of Marcet & Sargent (1989b), where type $K$ agent forms expectations according to

$$E^*_{Kt}(z_{1t}) = \Phi'_{Kt} z^K_{it-1} + \eta_t$$

where $z^K_{it}$ is possibly a subset of $z_{2t}$, i.e. the vector which contains exactly the variables that are relevant for predicting $z_{1t}$. With this setup, if convergence occurs then it will not be to 'standard' REEs, but to other equilibria which have appeared in the literature as *limited information rational expectations equilibria*, *restricted*

[8] A variant of this is the case where different groups of agents have different (mis)specifications of the model.

*perceptions equilibria* or *self-confirming equilibria*. In contrast, for the cases of heterogeneity analysed here, it is assumed that all groups of agents are aware of the correct specification of the model, but for various reasons their parameter estimates differ, i.e. $\{\Phi_{At}\}_{t=0}^{\infty} \neq \{\Phi_{Bt}\}_{t=0}^{\infty}$.

What follows is the description of the three types of heterogeneity and the corresponding results on stability conditions for each case.

**Agents with different expectations (or initial perceptions).** The first type of heterogeneity that is introduced in the model is a situation where the agents have different expectations about the economic variables, that is $\Phi_{At} \neq \Phi_{Bt}$ for at least some number of periods. In a setting where the agents are not fully rational, it is actually more reasonable to allow for this possibility than to assume that all agents have identical expectations. Implicitly, the latter assumption requires a great deal of expectational coordination in what the agents believe about the economy, which in turn hints at an underlying exogenous mechanism that dictates to the agents what precise expectation they should have. In contrast, allowing the agents to have different expectation can incorporate a situation where, due to psychological, cultural or other exogenous factors some agents are for example optimistic about the economy while others are pessimistic. The issue I wish to explore here is whether allowing the agents to have different expectations alters the evolution of the economic system in the sense of convergence to and stability of the REEs.

Formally, introducing heterogeneous expectations formation requires only that the agents have different initial beliefs about the parameters, i.e. that $\Phi_{A0} \neq \Phi_{B0}$. Assuming that the agents use recursive least squares to update their perceptions, the parameter estimates are updated according to

$$
\begin{aligned}
\Phi_{At} &= \Phi_{A,t-1} + \alpha_t R_{A,t-1}^{-1} z_{2t-1}\left[z_{2t-1}'\left(T\left(g\left(\mathbf{\Phi}_t\right)\right)' - \Phi_{A,t-1}'\right) + u_{t-1}' V\left(g\left(\mathbf{\Phi}_t\right)\right)'\right] \\
R_{At} &= R_{A,t-1} + \alpha_t\left[z_{2t-1}z_{2t-1}' - R_{At-1}/t\alpha_t\right] \\
\Phi_{Bt} &= \Phi_{B,t-1} + \alpha_t R_{B,t-1}^{-1} z_{2t-1}\left[z_{2t-1}'\left(T\left(g\left(\mathbf{\Phi}_t\right)\right)' - \Phi_{B,t-1}'\right) + u_{t-1}' V\left(g\left(\mathbf{\Phi}_t\right)\right)'\right] \\
R_{Bt} &= R_{B,t-1} + \alpha_t\left[z_{2t-1}z_{2t-1}' - R_{B,t-1}/t\alpha_t\right]
\end{aligned}
$$

while if they use stochastic gradient learning, the estimates are updated according to

$$
\begin{aligned}
\Phi_{At} &= \Phi_{A,t-1} + \alpha_t z_{2t-1}\left[z_{2t-1}'\left(T\left(g\left(\mathbf{\Phi}_t\right)\right)' - \Phi_{A,t-1}'\right) + u_{t-1}' V\left(g\left(\mathbf{\Phi}_t\right)\right)'\right] \\
\Phi_{Bt} &= \Phi_{B,t-1} + \alpha_t z_{2t-1}\left[z_{2t-1}'\left(T\left(g\left(\mathbf{\Phi}_t\right)\right)' - \Phi_{B,t-1}'\right) + u_{t-1}' V\left(g\left(\mathbf{\Phi}_t\right)\right)'\right]
\end{aligned}
$$

The following proposition determines the stability conditions for the above algorithms. Recall that $L(x) = d\ \mathrm{vec}(T(x)')\,/d\mathrm{vec}x$ and define the following 'weight' matrix

$$W = \begin{pmatrix} \psi & 1-\psi \\ \psi & 1-\psi \end{pmatrix}$$

**Proposition 4.1.** *When agents have different expectations about the parameters of the model and they update their perceptions using recursive least squares learning, the local asymptotic stability of an REE $\Phi_f$ is determined by the stability of the matrix*

$$J_1^{LS}(\Phi_f) = W \nwarrow\!\!\!\diagdown L(\Phi_f) - I_{2n_1n_2}$$

*This matrix is stable whenever $J^{LS}(\Phi_f)$ is stable. Furthermore, when the agents update their perceptions with stochastic gradient learning, the local asymptotic stability of $\Phi_f$ is determined by the stability of the matrix*

$$J_1^{SG}(\Phi_f) = (I_{2n_1} \nwarrow\!\!\!\diagdown M(\Phi_f)) \cdot J_1^{LS}(\Phi_f)$$

**Proof.** See appendix C. ■

This proposition suggests that differences in expectations do not matter when agents use least squares learning, or equivalently, that the stability conditions under homogeneous least squares learning (also known as E-stability conditions) are sufficient to ensure stability for this type of heterogeneity. Furthermore, when the agents use stochastic gradient learning and $n_1 = n_2 = 1$, it follows trivially that the E-stability conditions are sufficient for stability of $J_1^{SG}(\Phi_f)$. Although it is not in general true that if $J_1^{LS}(\Phi_f)$ is stable so is $J_1^{SG}(\Phi_f)$, it can be shown, as will be demonstrated in section 5, that for a number specific examples that cover a wide variety of economic models, it is indeed true.

**Agents with different degrees of inertia.** The second type of heterogeneity is a case where the agents have different degrees of inertia in their updating, in the sense of how much weight they put on the new incoming information in each period. The way an adaptive algorithm is interpreted is that in each period an agent updates the parameters of interest (here $\mathbf{\Phi}$) by adding to or substracting from his previous estimate a quantity which depends on the newly/currently observed information. Typically, this quantity reflects the forecasting error of the previous estimate. Furthermore, how important this quantity is for the agent is captured by the gain sequence $\{\alpha_t\}_{t=0}^{\infty}$ in the general algorithm. In essence, the absolute value of the gain sequence captures the degree of inertia of the agent in updating. There are several ways to introduce different degrees of inertia, which basically comes down to a variety of different gain sequences. Here, I analyse a simple case where the gain sequence of agent $B$ is a multiple or fraction of the gain sequence of agent $A$. Formally, for some $\delta > 0$, the two agents update the parameters according to the least squares learning algorithm

$$\begin{aligned}
\Phi_{At} &= \Phi_{A,t-1} + \alpha_t R_{A,t-1}^{-1} z_{2t-1} \left[ z_{2t-1}' \left( T\left(g\left(\mathbf{\Phi}_t\right)\right)' - \Phi_{A,t-1}' \right) + u_{t-1}' V\left(g\left(\mathbf{\Phi}_t\right)\right)' \right] \\
R_{At} &= R_{A,t-1} + \alpha_t \left[ z_{2t-1} z_{2t-1}' - R_{At-1}/t\alpha_t \right] \\
\Phi_{Bt} &= \Phi_{B,t-1} + \delta\alpha_t R_{B,t-1}^{-1} z_{2t-1} \left[ z_{2t-1}' \left( T\left(g\left(\mathbf{\Phi}_t\right)\right)' - \Phi_{B,t-1}' \right) + u_{t-1}' V\left(g\left(\mathbf{\Phi}_t\right)\right)' \right] \\
R_{Bt} &= R_{B,t-1} + \delta\alpha_t \left[ z_{2t-1} z_{2t-1}' - R_{B,t-1}/t\alpha_t \right]
\end{aligned}$$

The following proposition provides stability conditions for convergence to and stability of an REE for the above algorithm. Define the matrix $\Delta = diag\{1, \delta\}$.

**Proposition 4.2.** *When agents have different degrees of inertia in updating the parameters of the model and they update their perceptions using recursive least squares learning, the local asymptotic stability of an REE $\Phi_f$ is determined by the stability of the matrix*

$$J_2(\Phi_f) = \left(\Delta \nwarrow I_{n_1 n_2}\right) J_1^{LS}(\Phi_f)$$

**Proof.** See appendix D. ■

Unfortunately, it is not possible to show a general result by which $J_2(\Phi_f)$ is stable whenever the E-stability conditions are satisfied. However, here too, a great number of examples that have been examined indicate that this typically holds (at least for standard models). Some of these examples will be discussed in section 5. Besides, simple intuition suggests that, if some agents have more (or less) inertia than the rest of the population, this would at most lead to slower or faster adaptation, and hence a change in the rate of convergence, rather than preventing the algorithm from converging altogether.

**Agents that use different learning algorithms.** In the final case of heterogeneous learning, we let the agents use different learning algorithms. In particular, it is assumed that type $A$ agents update their perceptions using the recursive least squares algorithm, while type $B$ agents update their perceptions using the stochastic gradient algorithm, i.e. learning is occurring through the following mixed algorithm

$$\begin{aligned}
\Phi_{At} &= \Phi_{A,t-1} + \alpha_t R_{A,t-1}^{-1} z_{2t-1} \left[ z_{2t-1}' \left( T\left(g\left(\mathbf{\Phi}_t\right)\right)' - \Phi_{A,t-1}' \right) + u_{t-1}' V\left(g\left(\mathbf{\Phi}_t\right)\right)' \right] \\
R_{At} &= R_{A,t-1} + \alpha_t \left[ z_{2t-1} z_{2t-1}' - R_{At-1}/t\alpha_t \right] \\
\Phi_{Bt} &= \Phi_{B,t-1} + \alpha_t z_{2t-1} \left[ z_{2t-1}' \left( T\left(g\left(\mathbf{\Phi}_t\right)\right)' - \Phi_{B,t-1}' \right) + u_{t-1}' V\left(g\left(\mathbf{\Phi}_t\right)\right)' \right]
\end{aligned}$$

It is a well established fact that, although the two algorithms are quite similar, least squares is more efficient from an econometric view point, while stochastic gradient is less complex from a computational view point (as it does not involve the inversion of the second moment estimate $R$). Loosely speaking, this setup can be used to pin down the heterogeneity which is due to differences in the computational 'abilities' and 'capabilities' of the agents. For example, we could imagine that the least squares algorithm is used by agents that have access to powerful computational tools, such as computers, while on the other hand, the stochastic gradient algorithm is used by agents for whom it is very costly to perform complex calculations, and prefer to do less calculations than have high econometric efficiency. Stability conditions for an REE under the mixed algorithm are given in the following proposition:

**Proposition 4.3.** *When agents use different learning rules for updating the parameters of the model, namely recursive least squares and stochastic gradient learning, the local asymptotic stability of an REE $\Phi_f$ is determined by the stability of the matrix*

$$J_3(\Phi_f) = \begin{pmatrix} I_{n_1 n_2} & \mathbf{0} \\ \mathbf{0} & I_{n_1} \nwarrow M(\Phi_f) \end{pmatrix} J_1^{LS}(\Phi_f)$$

**Proof.** See appendix E ■

Once again, the proposition suggests that the E-stability conditions are not in general sufficient to ensure stability of the mixed algorithm of the general formulation of the model. However, for all the examples that I have examined, the E-stability conditions imply stability of $J_3(\Phi_f)$.

## 5. Examples

In this section I discuss some examples from the class self-referential linear stochastic models, and I apply the stability results derived in the previous section in order to examine the effects of allowing for heterogeneous learning on the stability of REEs. The examples analysed here have reduced forms with (i) date $t$ expectations of future variables (ii) date $t-1$ expectations of current variables, (iii) date $t-1$ expectations of current and future variables and finally (iv) lagged endogenous variables. For all the examples I concentrate only on the Minimal State Variable (henceforth MSV) solutions, which typically (but not always) correspond to the unique stationary solutions of the models. For all the examples, it is shown that the local stability of the REEs for the three cases of heterogeneity is determined by the E-stability conditions. The first three examples have a unique MSV rational expectations solution, while the last one can have multiple MSV REEs.

The choice of the models presented here is based on various factors. First, these models represent a good range of standard stochastic linear macroeconomic models; examples which can be expressed in these reduced forms are among others, the Cagan (1956) model of inflation, the Muth (1961) cobweb model, the Lucas (1973) island model, the Sargent & Wallace (1975) model, the Taylor (1977) real balance model, the Taylor (1980) model of overlapping wage contracts, as well as several multivariate linear models, including log-linearisations of real business cycles models. For discussions of these examples and how they fit in the corresponding reduced forms, see Evans & Honkapohja (2001). Second, each model has a particular structural characteristic that makes the technical analysis interesting. Last, for the illustrational purposes of this section, the simplicity of the models allows for a straightforward analysis and conveys some clear messages, without having to engage in long algebraic calculations.

**Models with date $t$ expectations of future variables.** Consider a model that can be written in the reduced form

$$\begin{aligned} y_t &= \lambda E_t^* y_{t+1} + \kappa w_t \\ w_t &= \rho w_{t-1} + u_t \end{aligned}$$

where $\{w_t\}$ is an AR(1) exogenous variable with $u_t \sim (0, \sigma_u^2)$. Assuming that the representative agent forms expectations according to $E_{t-1}^* y_t = \phi_{t-1} w_{t-1}$, it follows that $T(\phi) = (\lambda\phi + 1)\rho$, hence $L(\phi) = \lambda\rho$. The unique fixed point of the $T-$ map is $\phi_f = \rho/(1 - \lambda\rho)$. The E-stability condition which is sufficient for stability of the REE under homogeneous least squares learning is that $\lambda\rho < 1$. Furthermore, the second moment of $z_{2t} = w_t$ is $M = \sigma^2 = \sigma_u^2/(1 - \rho^2)$. The matrices that determine the stability of the REE under the three types of heterogeneity are[9]

$$\begin{aligned} J_1^{LS}(\phi_f) &= \begin{pmatrix} \psi\lambda\rho - 1 & (1-\psi)\lambda\rho \\ \psi\lambda\rho & (1-\psi)\lambda\rho - 1 \end{pmatrix} \\ J_1^{SG}(\phi_f) &= \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \cdot J_1^{LS}(\phi_f) = \sigma^2 J_1^{LS}(\phi_f) \\ J_2(\phi_f) &= \begin{pmatrix} \psi\lambda\rho - 1 & (1-\psi)\lambda\rho \\ \delta\psi\lambda\rho & \delta\left[(1-\psi)\lambda\rho - 1\right] \end{pmatrix} \\ J_3(\phi_f) &= \begin{pmatrix} \psi\lambda\rho - 1 & (1-\psi)\lambda\rho \\ \sigma^2\psi\lambda\rho & \sigma^2\left[(1-\psi)\lambda\rho - 1\right] \end{pmatrix} \end{aligned}$$

[9] For this first example the relevant matrices are stated explicitly for illustrational purposes, but will be omitted for the rest of the examples, as their derivation is a straightforward algebraic exercise.

Proposition 4.1 ensures that $J_1^{LS}(\phi_f)$ is stable as long as $\lambda\rho < 1$. The same is trivially true for $J_1^{SG}(\phi_f)$, since $\sigma^2 > 0$. Furthermore, the eigenvalues of $J_2(\phi_f)$ are

$$\frac{1}{2}\left[\psi\lambda\rho - 1 + \delta\left((1-\psi)\lambda\rho - 1\right) \pm \sqrt{4\delta\left(\lambda\rho - 1\right) + \left(\psi\lambda\rho - 1 + \delta\left((1-\psi)\lambda\rho - 1\right)\right)^2}\right]$$

which can easily be shown to be negative if $\lambda\rho < 1$. With the same argument it follows that $J_3(\phi_f)$ is stable if the E-stability condition holds.

Examples of models that can be written in the above reduced form are the Cagan (1956) model of inflation, and an asset pricing model with risk neutrality, where the price of an asset at time $t$ is given by the rule

$$p_t = (1+r)^{-1}\left(E_t^* p_{t+1} + d_t\right)$$

where $r$ is the interest rate, and $d_t$ is the dividend the asset pays at the end of period $t$.

**Models with date $t-1$ expectations of current variables.** Suppose now that the model can be written in the following reduced form

$$\begin{aligned} p_t &= \mu + \alpha E_{t-1}^* p_t + \gamma w_t \\ w_t &= \kappa + \rho w_{t-1} + u_t \end{aligned}$$

where $\{w_t\}$ is an AR(1) exogenous variable with $u_t \sim (0, \sigma_u^2)$. If the representative agent form expectations according to $E_{t-1}^* y_t = a_{t-1} + b_{t-1} w_{t-1} \equiv \Phi_{t-1}' z_{2t-1}$, where $z_{2t-1} = (1, w_{t-1})$, it follows that

$$T(\Phi) = T\left((a,b)'\right) = \left(\begin{array}{cc} \mu + \gamma\kappa + \alpha a & \gamma\rho + \alpha b \end{array}\right)$$

and therefore $L(\Phi) = diag\{\alpha, \alpha\} = \alpha I_2$. The unique fixed point of the $T-$map is $\Phi_f = \left((1-\alpha)^{-1}(\mu + \gamma\kappa), (1-\alpha)^{-1}\gamma\rho\right)'$. The E-stability condition is now $\alpha < 1$. Let $m = \kappa/(1-\beta)$ and $\sigma^2 = \sigma_u^2/(1-\rho^2)$. The second moment matrix of $z_{2t}$ is then

$$M = \left(\begin{array}{cc} 1 & m \\ m & m^2 + \sigma^2 \end{array}\right)$$

The matrix $J_1^{LS}(\Phi_f)$ which determines the local asymptotic stability of the REE for the heterogeneous expectations least squares algorithm, is stable when $\alpha < 1$. Furthermore, $J_1^{SG}(\Phi_f)$ is also stable when $\alpha < 1$. This is because[10]

[10]Derivation:

$$\begin{aligned} J_1^{LS}(\Phi_f) &= (I_2 \nwarrow M)(\alpha W \nwarrow I_2 - I_4) = (I_2 \nwarrow M)(\alpha W \nwarrow I_2) - (I_2 \nwarrow M) \\ &= (\alpha W \nwarrow M) - (I_2 \nwarrow M) = (\alpha W - I_2) \nwarrow M \end{aligned}$$

$J_1^{LS}(\Phi_f) = (\alpha W - I_2) \nwarrow\!\!\searrow M$ and its eigenvalues are the products of the eigenvalues of $M$, which are always positive, and the eigenvalues of $\alpha W - I_2$ which are $-1$ and $\alpha - 1$, which are both negative as long as $\alpha < 1$.

Furthermore, for the case of different degrees of inertia, the eigenvalues of $J_2(\Phi_f)$ are

$$\frac{1}{2}\left[\psi\alpha - 1 + \delta\left((1-\psi)\alpha - 1\right) \pm \sqrt{4\delta\left(\alpha - 1\right) + \left(\psi\alpha - 1 + \delta\left((1-\psi)\alpha - 1\right)\right)^2}\right]$$

which are negative as long as $\alpha < 1$.

Last, for the mixed algorithm, although the eigenvalues of $J_3(\Phi_f)$ are too lengthy to appear here, it can be verified that they are real and negative.

Examples of models that can be written in this reduced form include the Muth (1961) cobweb model, and the Lucas (1973) island model.

**Models with date $t-1$ expectations of current and future variables.** Consider now a model that can be written in the reduced form

$$\begin{aligned} y_t &= \mu + \alpha E^*_{t-1} y_t + \beta E^*_{t-1} y_{t+1} + \gamma w_t \\ w_t &= \rho w_{t-1} + u_t \end{aligned}$$

where $\{w_t\}$ is an AR(1) exogenous variable with $u_t \sim (0, \sigma_u^2)$. Models of this form exhibit MSV solutions, as well as a continuum of, possibly stationary, sunspot/bubble RE solutions. Here I concentrate on the MSV solutions, as it is not possible to study analytically the stability of real time learning for multiple REEs that are not discrete. For this class of solutions, the representative agent's perceptions are formed according to $E^*_{t-1} y_t = a_{t-1} + b_{t-1} w_{t-1} \equiv \Phi'_{t-1} z_{2t-1}$, where $z_{2t-1} = (1, w_{t-1})$. It follows that

$$T(\Phi) = T\left((a,b)'\right) = \left(\begin{array}{cc} \mu + (\alpha+\beta)a, & (\alpha+\beta\rho)b + \gamma\rho \end{array}\right)$$

hence $L(\Phi) = diag\{\alpha+\beta, \alpha+\beta\rho\}$. The fixed point of the $T-$map is $\Phi_f = \left(\mu(1-\alpha-\beta)^{-1}, \gamma\rho(1-\alpha-\beta)^{-1}\right)'$. The second moment matrix of $z_{2t-1}$ is $M = diag\{1, \sigma^2\}$, where $\sigma^2 = \sigma_u^2/(1-\rho^2)$.

The matrix $J_1^{LS}(\Phi_f)$ which determines the local asymptotic stability of the REE for the heterogeneous expectations least squares algorithm is stable when the E-stability conditions hold, i.e. $\alpha+\beta < 1$ and $\alpha+\beta\rho < 1$. The matrix $J_1^{SG}(\Phi_f)$ has eigenvalues $-1, -\sigma^2, \alpha+\beta-1$, and $\alpha+\beta\rho-1$ which are negative under the same conditions.

Furthermore, for the case of different degrees of inertia, the eigenvalues of $J_2(\Phi_f)$ are $\frac{1}{2}\left[C \pm \sqrt{4\delta\left(\alpha+\beta-1\right) + C^2}\right]$ and $\frac{1}{2}\left[G \pm \sqrt{4\delta\left(\alpha+\beta\rho-1\right) + G^2}\right]$ where

$$\begin{aligned} C &= \psi\left(\alpha+\beta\right) - 1 + \delta\left[(1-\psi)\left(\alpha+\beta\right) - 1\right] \\ G &= \psi\left(\alpha+\beta\rho\right) - 1 + \delta\left[(1-\psi)\left(\alpha+\beta\rho\right) - 1\right] \end{aligned}$$

These are negative provided that the same conditions hold.

Finally, for the mixed algorithm, the eigenvalues of $J_3(\Phi_f)$ are $-1$, $\alpha+\beta-1$, and $\left[F \pm \sqrt{4\delta(\alpha+\beta\rho-1)+F^2}\right]/2$ where

$$F = \psi(\alpha+\beta\rho) - 1 + \sigma^2((1-\psi)(\alpha+\beta\rho)-1)$$

These eigenvalues are also negative under the same conditions.

Examples of models that can be written in this reduced form include the Sargent & Wallace (1975) model, and the Taylor (1977) real balance model.

**Models with lagged endogenous variables.** Finally consider a model that can be written in a reduced form that contains lags of the endogenous variables. Suppose that we can write the model as

$$y_t = \lambda y_{t-1} + \alpha E^*_{t-1} y_t + \beta E^*_{t-1} y_{t+1} + u_t$$

where $u_t$ is a $(0, \sigma^2)$ error term. The perceptions of the representative agent evolve according to $E^*_{t-1} y_t = \phi_{t-1} y_{t-1}$. Substituting this back to the reduced form of the model we find that $T(\phi) = \lambda + \alpha\phi + \beta\phi^2$. This mapping has two real fixed points (REEs) provided that $D = (\alpha-1)^2 - 4\beta\lambda > 0$, which are stationary if they are smaller than one in absolute value. If these conditions are satisfied then the REEs are

$$\bar{\phi}_{1,2} = \frac{1}{2\beta}\left(1 - \alpha \pm \sqrt{D}\right)$$

The second moment matrix of $z_{2t} = y_{t-1}$ is $M(\phi) = \sigma^2/\left(1 - T(\phi)^2\right)$. Furthermore, $L(\phi) = \alpha + 2\beta\phi$. Under homogeneous learning (both using least squares and stochastic gradient algorithms) the first REE is never stable. This is because $L(\bar{\phi}_1) - 1 = \sqrt{D} > 0$. On the other hand, the second REE is always stable since $L(\bar{\phi}_2) - 1 = -\sqrt{D} < 0$.

The stability properties of the two REEs are preserved locally for the case of heterogeneous expectations, both for least squares and stochastic gradient learning. For stochastic gradient learning with heterogeneous expectations, $J_1^{SG}(\phi_f) = M(\phi_f) \cdot J_1^{LS}(\phi_f)$, where $M(\phi_f)$ is a positive scalar. Therefore the signs of the eigenvalues of $J_1^{SG}(\phi_f)$ are the same as the signs of the eigenvalues of $J_1^{LS}(\phi_f)$.

For the case of agents with different degrees of inertia, the eigenvalues of $J_2(\bar{\phi}_1)$ are $\frac{1}{2}\left[K \pm \sqrt{4\delta\sqrt{D} + K^2}\right]$ where

$$K = \delta\left(\sqrt{D}(1-\psi) - \psi\right) + \psi\sqrt{D} - (1-\psi)$$

The large eigenvalue is always positive, and therefore $\bar{\phi}_1$ is unstable.

Furthermore the eigenvalues of $J_2(\bar{\phi}_2)$ are $-\frac{1}{2}\left[L \pm \sqrt{-4\delta\sqrt{D}+L^2}\right]$ where

$$L = \delta\left(\sqrt{D}(1-\psi)+\psi\right)+\psi\sqrt{D}+(1-\psi)$$

Both the eigenvalues are always negative, hence $\bar{\phi}_2$ is stable.

For the case of the mixed algorithm, the stability properties are again preserved, since the eigenvalues of $J_3(\bar{\phi}_i)$ are the same as the eigenvalues of $J_2(\bar{\phi}_i)$ after substituting $M(\bar{\phi}_i)$ for $\delta$.

Examples of models that can be written in this reduced form include the special case of a two period Taylor (1980) overlapping wage contract model, and the Taylor (1977) model augmented with a policy feedback rule.

## 6. Closing comments

Although the analysis presented here does not claim to be exhaustive, it provides a step towards a better understanding of how heterogeneity might affect learning. The general formulation analysed here covers a very wide range of macroeconomic models, which, apart from standard univariate cases, includes linearisations of multivariate models, such as real business cycle models. The fact that for this class of models it cannot be shown that the stability conditions for heterogeneous learning are the same as the ones for the homogeneous case could be alarming news for proponents of the representative agent. But as demonstrated by the examples, it appears that it is often the case that aggregating is safe. The point I wish to stress, based on the present results, is that the representative agent is (perhaps surprisingly) often a good approximation of the agents in an economy, but any rigorous analysis should include a test of the assumption, for example a test along the lines suggested here.

Initiating from the present analysis, there are several further issues worthy of further exploration. For example, the results presented here leave out any inference on the global dynamics of the system under heterogeneous expectations. Preliminary numerical investigation of the global behaviour of examples that exhibit multiple REEs indicates that the representative agent is indeed a very good approximation, yet a rigorous argument still remains unavailable. Furthermore, another important aspect besides the stability of an REE is the rate with which the learning algorithm converges to it. Numerical estimation of the rates of convergence for the stochastic cobweb model (second example in section 5) with heterogeneity (see Giannitsarou (2001)) gives strong evidence that the rates can be very different from and often much higher than the corresponding homogeneous case. Both the issues of global stability and the rates of convergence are important in models where we are interested in the off-equilibrium dynamics, such as models that study the effects of monetary or fiscal reforms, financial asset pricing models, or exchange rate models.

Finally, it would be interesting to find a model for which the representative agent is not a good approximation, in the sense that further conditions are required to ensure stability of the REEs under heterogeneous learning. Exploring what the driving force for the differentiation between the representative and the heterogeneous agents is, could provide very useful insights about how heterogeneity matters, if it does matter at all.

References


1. Barucci E., 1999. Heterogeneous Beliefs and Learning in Forward Looking Economic Models. *Evolutionary Economics*, 9, 453 - 464.

2. Barucci, E. and L. Landi, 1997. Least Mean Squares Learning in Self-referential Linear Stochastic Models. *Economics Letters* 57, 313 - 317.

3. Cagan, P., 1956. The Monetary Dynamics of Hyper-Inflation. In 'Studies in the Quantity Theory of Money', ed. by M. Friedman, *University of Chicago Press,* Chicago.

4. Bray, M. M. and N. Savin, 1986. Rational Expectations Equilibria, Learning and Model Specification. *Econometrica* 54, 1129 - 1160.

5. Evans, G. and R. Guesnerie, 1999. Coordination on Saddle Path Solutions: the Eductive Viewpoint. 1-Linear Univariate models. *Working paper.*

6. Evans, G. and S. Honkapohja, 1996. Least Squares Learning with Heterogeneous Expectations. *Economics Letters* 53, 197 - 201.

7. Evans, G. and S. Honkapohja, 1998a. Convergence of Learning algorithms without a Projection Facility. *Journal of Mathematical Economics,* 30, 59 - 86.

8. Evans, G. and S. Honkapohja, 1998b. Stochastic Gradient Learning in the Cobweb Model. *Economics Letters* 61, 333 - 337.

9. Evans, G. and S. Honkapohja, 2001. Learning and Expectations in Macroeconomics. *Princeton University Press.*

10. Evans, G., S. Honkapohja and R. Marimon, 2000. Convergence in Monetary Inflation Models with Heterogeneous Learning Rules. *Macroeconomic Dynamics,* forthcoming.

11. Franke, R. and T. Nesemann, 1999. Two Destabilizing Strategies May Be Jointly Stabilizing. *Journal of Economics*, 1, 1 - 18.

12. Giannitsarou, C., 2001. Rates of Convergence of Learning with Heterogeneity in the Stochastic Cobweb Model. *Mimeo* in progress.

13. Hahn, F. and R. Solow, 1997. A Critical Essay on Modern Macroeconomic Theory. *Blackwell Publishers*, Oxford.

14. Heinemann, M. 2000. Convergence of Adaptive Learning and Expectational Stability: the Case of Multiple Rational-Expectations Equilibria. *Macroeconomic Dynamics*, 4 (3).

15. Kuan, C.-M. and H. White, 1994. Adaptive Learning with Nonlinear Dynamics Driven by Dependent Processes. *Econometrica*, 62, 1087 - 1114.

16. Ljung, L., 1977. Analysis of Recursive Stochastic Algorithms. *IIIE Transactions on Automatic Control*, AC - 22, 551 - 575.

17. Lucas, R. 1973. Some International Evidence on Output - Inflation Tradeoffs. *American Economic Review*, 63, 326 - 334.

18. Magnus, J. and H. Neudecker, 1988. Matrix Differential Calculus. *Wiley*, New York.

19. Marcet, A. and T. Sargent, 1989a. Convergence of Least Squares Mechanisms in Self-Referential Linear Stochastic Models. *Journal of Economic Theory* 48, 337 - 368.

20. Marcet, A. and T. Sargent, 1989b. Convergence of Least Squares Learning in Environments with Hidden State Variables and Private Information. *Journal of Political Economy* 97, 1306 - 1322.

21. Muth, J. 1961. Rational Expectations and the Theory of Price Movements. *Econometrica*, 29, 315 - 335.

22. Sargent, T., 1993. Bounded Rationality in Macroeconomics. *Oxford University Press*, Oxford.

23. Sargent, T. and N. Wallace, 1975. 'Rational Expectations', the Optimal Monetary Instrument and the Optimal Money Supply Rule. *Journal of Political Economy*, 83, 241-254.

24. Taylor, J., 1977. Conditions for Unique Solutions in Stochastic Macroeconomic Models with Rational Expectations. *Econometrica*, 45, 1377 - 1386.

25. Taylor, J., 1980. Aggregate Dynamics and Staggered Contracts. *Journal of Political Economy*, 88, 1 - 23.

Appendices

*The results on matrix differential calculus that have been used in the following appendices are taken from Magnus & Neudecker (1988).*

## A. Technical assumptions for the ode method

- **A1.** $\alpha_t > 0$ for all $t$, is a deterministic, non-increasing sequence such that $\sum_{t=1}^{\infty} \alpha_t = \infty$ and $\sum_{t=1}^{\infty} \alpha_t^2 < \infty$.
- **A2.** For any compact set $H \subset D$ there exist $C$ and $q$ such that $|Q(\theta, z)| \leq C(1 + |z|^q)$ for all $\theta \in H$ and for all $t$.
- **A3.** For any compact set $H \subset D$ and for all $\theta, \theta' \in H$ and $z_1, z_2 \in \mathbb{R}^k$, the function $Q(\theta, z)$ satisfies:
  1. $|\partial Q(\theta, z_1)/\partial z - \partial Q(\theta, z_2)/\partial z| \leq L_1 |z_1 - z_2|$
  2. $|Q(\theta, 0) - Q(\theta', 0)| \leq L_2 |\theta - \theta'|$
  3. $|\partial Q(\theta, z)/\partial z - \partial Q(\theta', z)/\partial z| \leq L_2 |\theta - \theta'|$
- **B1.** $\mathbf{u}_t$ is iid with finite absolute moments.
- **B2.** For any compact set $H \subset D$, $\sup_{\theta \in H} |C(\theta)| \leq M$ and $\sup_{\theta \in H} |G(\theta)| < 1$ where $C(.)$ and $G(.)$ are defined by the expression

$$z_t = G(\theta_{t-1}) z_{t-1} + C(\theta_{t-1}) \mathbf{u}_t$$

## B. Derivation of $J^{SG}(\Phi)$

The Jacobian for homogeneous stochastic gradient learning is

$$J^{SG}(\Phi) = \frac{d \text{ vec}\left[M(\Phi) \cdot (T(\Phi)' - \Phi)\right]}{d \text{ vec}\Phi}$$

First note that

$$\begin{aligned}
d \text{ vec}\left[M(\Phi) \cdot (T(\Phi)' - \Phi)\right] &= \\
\text{vec } d\left[M(\Phi) \cdot (T(\Phi)' - \Phi)\right] &= \\
\text{vec } \left[(dM(\Phi)) \cdot (T(\Phi)' - \Phi) + M(\Phi) \cdot d\,(T(\Phi)' - \Phi)\right] &= \\
\text{vec } \left[(dM(\Phi)) \cdot (T(\Phi)' - \Phi)\right] + \text{vec } \left[M(\Phi) \cdot d\,(T(\Phi)' - \Phi) \cdot I\right] &= \\
\left[(T(\Phi)' - \Phi) \nwarrow I\right] \cdot \text{vec}dM(\Phi) + (I \nwarrow M(\Phi)) \cdot \text{vec } d\,(T(\Phi)' - \Phi) &= \\
\left[(T(\Phi)' - \Phi) \nwarrow I\right] \cdot d\text{vec}M(\Phi) + (I \nwarrow M(\Phi)) \cdot d\text{vec}\,(T(\Phi)' - \Phi) &=
\end{aligned}$$

and therefore

$$
\begin{aligned}
J^{SG}(\Phi) &= \frac{d \ \mathrm{vec}\left[M(\Phi)\left(T(\Phi)' - \Phi\right)\right]}{d \ \mathrm{vec}\Phi} \\
&= \left[\left(T(\Phi)' - \Phi\right) \otimes I\right] \cdot \frac{d\mathrm{vec}M(\Phi)}{d \ \mathrm{vec}\Phi} + \left(I \otimes M(\Phi)\right) \cdot \frac{d\mathrm{vec}\left(T(\Phi)' - \Phi\right)}{d \ \mathrm{vec}\Phi} \\
&= \left[\left(T(\Phi)' - \Phi\right) \otimes I\right] \cdot \frac{d\mathrm{vec}M(\Phi)}{d \ \mathrm{vec}\Phi} + \left(I \otimes M(\Phi)\right) \cdot J^{LS}(\Phi)
\end{aligned}
$$

Furthermore, the Jacobian evaluated at $\Phi_f$ is $(I \otimes M(\Phi_f)) \cdot J^{LS}(\Phi_f)$, since $T(\Phi_f)' = \Phi_f$ ■

# C. Proof of proposition 4.1

The least squares algorithm for heterogeneous expectations can be associated to the big ode

$$
\begin{aligned}
\frac{d\Phi_A}{d\tau} &= R_A^{-1} M\left(g\left(\mathbf{\Phi}\right)\right)\left[T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_A\right] \\
\frac{dR_A}{d\tau} &= M\left(g\left(\mathbf{\Phi}\right)\right) - R_A \\
\frac{d\Phi_B}{d\tau} &= R_B^{-1} M\left(g\left(\mathbf{\Phi}\right)\right)\left[T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_B\right] \\
\frac{dR_B}{d\tau} &= M\left(g\left(\mathbf{\Phi}\right)\right) - R_B
\end{aligned}
$$

The local stability of an REE $\Phi_f$ is therefore determined by the vectorised version of the small ode

$$
\begin{aligned}
\frac{d\mathbf{\Phi}}{d\tau} &= \begin{pmatrix} \frac{d\Phi_A}{d\tau}, & \frac{d\Phi_B}{d\tau} \end{pmatrix} \\
&= \begin{pmatrix} T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_A, & T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_B \end{pmatrix}
\end{aligned}
$$

Therefore the relevant Jacobian is

$$
\begin{aligned}
J_1^{LS}\left(\Phi_f\right) &= \frac{d}{d \ \mathrm{vec}\mathbf{\Phi}} \begin{pmatrix} \mathrm{vec} \ \left[T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_A\right] \\ \mathrm{vec} \ \left[T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_B\right] \end{pmatrix}\Bigg|_{\mathbf{\Phi}=\left(\Phi_f, \Phi_f\right)} \\
&= \begin{pmatrix} \frac{d \ \mathrm{vec}\left[T(g(\mathbf{\Phi}))' - \Phi_A\right]}{d \ \mathrm{vec}\Phi_A} & \frac{d \ \mathrm{vec}\left[T(g(\mathbf{\Phi}))' - \Phi_A\right]}{d \ \mathrm{vec}\Phi_B} \\ \frac{d \ \mathrm{vec}\left[T(g(\mathbf{\Phi}))' - \Phi_B\right]}{d \ \mathrm{vec}\Phi_A} & \frac{d \ \mathrm{vec}\left[T(g(\mathbf{\Phi}))' - \Phi_B\right]}{d \ \mathrm{vec}\Phi_B} \end{pmatrix}\Bigg|_{\mathbf{\Phi}=\left(\Phi_f, \Phi_f\right)} \\
&= \begin{pmatrix} \frac{d \ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d \ \mathrm{vec}\Phi_A} - I_{n_1 n_2} & \frac{d \ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d \ \mathrm{vec}\Phi_B} \\ \frac{d \ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d \ \mathrm{vec}\Phi_A} & \frac{d \ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d \ \mathrm{vec}\Phi_B} - I_{n_1 n_2} \end{pmatrix}\Bigg|_{\mathbf{\Phi}=\left(\Phi_f, \Phi_f\right)}
\end{aligned}
$$

Applying the chain rule for differentiating vectors we obtain that

$$\left.\frac{d\ \mathrm{vec}T\left(g\left(\mathbf{\Phi}\right)\right)'}{d\ \mathrm{vec}\Phi_A}\right|_{\mathbf{\Phi}=\left(\Phi_f,\Phi_f\right)}=\left.\frac{d\ \mathrm{vec}T\left(\Phi\right)'}{d\ \mathrm{vec}\Phi}\right|_{\Phi=\Phi_f}\cdot\left.\frac{d\ \mathrm{vec}g\left(\mathbf{\Phi}\right)}{d\ \mathrm{vec}\Phi_A}\right|_{\mathbf{\Phi}=\left(\Phi_f,\Phi_f\right)}=\psi L(\Phi_f)$$

Similarly

$$\frac{d\ \mathrm{vec}T\left(g\left(\mathbf{\Phi}\right)\right)'}{d\ \mathrm{vec}\Phi_B}=(1-\psi)L(\Phi_f)$$

Hence

$$\begin{aligned} J_1^{LS}(\Phi_f) &= \begin{pmatrix} \psi L(\Phi_f)-I_{n_1n_2} & (1-\psi)L(\Phi_f) \\ \psi L(\Phi_f) & (1-\psi)L(\Phi_f)-I_{n_1n_2} \end{pmatrix} \\ &= \begin{pmatrix} \psi L(\Phi_f) & (1-\psi)L(\Phi_f) \\ \psi L(\Phi_f) & (1-\psi)L(\Phi_f) \end{pmatrix} - I_{2n_1n_2} \\ &= W \nwarrow\!\!\searrow L(\Phi_f)-I_{2n_1n_2} \end{aligned}$$

where

$$W=\begin{pmatrix} \psi & 1-\psi \\ \psi & 1-\psi \end{pmatrix}$$

Let $\lambda_i$ be the eigenvalues of $L(\Phi_f)$. To see why this matrix is stable whenever $J^{LS}(\Phi_f)$ is stable, note that if $\Phi_f$ is locally stable under the homogeneous least squares algorithm, all the eigenvalues of the matrix $J(\Phi_f)=L(\Phi_f)-I_{n_1n_2}$ have negative real parts, i.e. that $\mathrm{Re}(\lambda_i)<1$, for all $i=1,...,n_1n_2$. Furthermore, the eigenvalues of $W$ are 0 and 1, therefore the eigenvalues of $W \nwarrow\!\!\searrow L(\Phi_f)$ are 0 (with multiplicity $n_1n_2$) and $\lambda_i$ and it follows that the eigenvalues of $J_1(\Phi_f)$ have real parts $-1<0$ or $\mathrm{Re}(\lambda_i)-1<0$. Hence $J_1(\Phi_f)$ is stable.

For the second part of the proposition, the stochastic gradient algorithm for heterogeneous expectations can be associated to the ode

$$\begin{aligned} \frac{d\Phi_A}{d\tau} &= M\left(g\left(\mathbf{\Phi}\right)\right)\left[T\left(g\left(\mathbf{\Phi}\right)\right)'-\Phi_A\right] \\ \frac{d\Phi_B}{d\tau} &= M\left(g\left(\mathbf{\Phi}\right)\right)\left[T\left(g\left(\mathbf{\Phi}\right)\right)'-\Phi_B\right] \end{aligned}$$

The local stability of an REE $\Phi_f$ is therefore determined by the vectorised version of the small ode

$$\begin{aligned} \frac{d\mathbf{\Phi}}{d\tau} &= \left(\ \tfrac{d\Phi_A}{d\tau},\ \ \tfrac{d\Phi_B}{d\tau}\ \right) \\ &= \left(\ M\left(g\left(\mathbf{\Phi}\right)\right)\left[T\left(g\left(\mathbf{\Phi}\right)\right)'-\Phi_A\right],\ \ M\left(g\left(\mathbf{\Phi}\right)\right)\left[T\left(g\left(\mathbf{\Phi}\right)\right)'-\Phi_B\right]\ \right) \end{aligned}$$

and the corresponding Jacobian of the vectorised ode

$$J_1^{SG}(\Phi_f) = \frac{d}{d\ \mathrm{vec}\mathbf{\Phi}} \begin{pmatrix} \mathrm{vec}\ \left\{M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))' - \Phi_A\right]\right\} \\ \mathrm{vec}\ \left\{M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))' - \Phi_B\right]\right\} \end{pmatrix}\Bigg|_{\mathbf{\Phi}=(\Phi_f,\Phi_f)}$$

$$= \begin{pmatrix} \frac{d\ \mathrm{vec}\left\{M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))'-\Phi_A\right]\right\}}{d\ \mathrm{vec}\Phi_A} & \frac{d\ \mathrm{vec}\left\{M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))'-\Phi_A\right]\right\}}{d\ \mathrm{vec}\Phi_B} \\ \frac{d\ \mathrm{vec}\left\{M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))'-\Phi_B\right]\right\}}{d\ \mathrm{vec}\Phi_A} & \frac{d\ \mathrm{vec}\left\{M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))'-\Phi_B\right]\right\}}{d\ \mathrm{vec}\Phi_B} \end{pmatrix}\Bigg|_{\mathbf{\Phi}=(\Phi_f,\Phi_f)}$$

Using similar arguments as in appendix A, it follows that

$$\begin{aligned} J_1^{SG}(\Phi_f) &= \begin{pmatrix} [I_{n_1} \nwarrow M(\Phi_f)]\,[\psi L(\Phi_f) - I_{n_1 n_2}] & [I_{n_1} \nwarrow M(\Phi_f)]\,(1-\psi)L(\Phi_f) \\ [I_{n_1} \nwarrow M(\Phi_f)]\,\psi L(\Phi_f) & [I_{n_1} \nwarrow M(\Phi_f)]\,[(1-\psi)L(\Phi_f) - I_{n_1 n_2}] \end{pmatrix} \\ &= \begin{pmatrix} I_{n_1} \nwarrow M(\Phi_f) & \mathbf{0} \\ \mathbf{0} & I_{n_1} \nwarrow M(\Phi_f) \end{pmatrix} \begin{pmatrix} \psi L(\Phi_f) - I_{n_1 n_2} & (1-\psi)L(\Phi_f) \\ \psi L(\Phi_f) & (1-\psi)L(\Phi_f) - I_{n_1 n_2} \end{pmatrix} \\ &= (I_{2n_1} \nwarrow M(\Phi_f))\, J_1^{LS}(\Phi_f) \end{aligned}$$

■

# D. Proof of proposition 4.2

The algorithm for different degrees of inertia can be associated to the big ode

$$\begin{aligned} \frac{d\Phi_A}{d\tau} &= R_A^{-1} M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))' - \Phi_A\right] \\ \frac{dR_A}{d\tau} &= M(g(\mathbf{\Phi})) - R_A \\ \frac{d\Phi_B}{d\tau} &= \delta R_B^{-1} M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))' - \Phi_B\right] \\ \frac{dR_B}{d\tau} &= \delta\left[M(g(\mathbf{\Phi})) - R_B\right] \end{aligned}$$

The local stability of an REE $\Phi_f$ is therefore determined by the vectorised version of the small ode

$$\frac{d\mathbf{\Phi}}{d\tau} = \left( \frac{d\Phi_A}{d\tau},\ \ \frac{d\Phi_B}{d\tau} \right) = \left( T(g(\mathbf{\Phi}))' - \Phi_A,\ \ \delta\left[T(g(\mathbf{\Phi}))' - \Phi_B\right] \right)$$

Therefore the relevant Jacobian is

$$J_2(\Phi_f) = \frac{d}{d\ \mathrm{vec}\mathbf{\Phi}} \begin{pmatrix} \mathrm{vec}\ \left[T(g(\mathbf{\Phi}))' - \Phi_A\right] \\ \mathrm{vec}\ \delta\left[T(g(\mathbf{\Phi}))' - \Phi_B\right] \end{pmatrix}\Bigg|_{\mathbf{\Phi}=(\Phi_f,\Phi_f)}$$

$$
\begin{aligned}
&= \left( \begin{array}{cc} \frac{d\ \mathrm{vec}\left[T(g(\mathbf{\Phi}))'-\Phi_A\right]}{d\ \mathrm{vec}\Phi_A} & \frac{d\ \mathrm{vec}\left[T(g(\mathbf{\Phi}))'-\Phi_A\right]}{d\ \mathrm{vec}\Phi_B} \\ \delta\frac{d\ \mathrm{vec}\left[T(g(\mathbf{\Phi}))'-\Phi_B\right]}{d\ \mathrm{vec}\Phi_A} & \delta\frac{d\ \mathrm{vec}\left[T(g(\mathbf{\Phi}))'-\Phi_B\right]}{d\ \mathrm{vec}\Phi_B} \end{array} \right)\Bigg|_{\mathbf{\Phi}=\left(\Phi_f,\Phi_f\right)} \\
&= \left( \begin{array}{cc} \frac{d\ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d\ \mathrm{vec}\Phi_A} - I_{n_1 n_2} & \frac{d\ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d\ \mathrm{vec}\Phi_B} \\ \delta\frac{d\ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d\ \mathrm{vec}\Phi_A} & \delta\frac{d\ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d\ \mathrm{vec}\Phi_B} - \delta I_{n_1 n_2} \end{array} \right)\Bigg|_{\mathbf{\Phi}=\left(\Phi_f,\Phi_f\right)} \\
&= \left( \begin{array}{cc} I_{n_1 n_2} & \mathbf{0} \\ \mathbf{0} & \delta I_{n_1 n_2} \end{array} \right) \left( \begin{array}{cc} \frac{d\ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d\ \mathrm{vec}\Phi_A} - I_{n_1 n_2} & \frac{d\ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d\ \mathrm{vec}\Phi_B} \\ \frac{d\ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d\ \mathrm{vec}\Phi_A} & \frac{d\ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d\ \mathrm{vec}\Phi_B} - I_{n_1 n_2} \end{array} \right)\Bigg|_{\mathbf{\Phi}=\left(\Phi_f,\Phi_f\right)} \\
&= \left(\Delta \nwarrow I_{n_1 n_2}\right) J_1^{LS}(\Phi_f)
\end{aligned}
$$

where

$$
\Delta = \left( \begin{array}{cc} 1 & 0 \\ 0 & \delta \end{array} \right)
$$

■

# E. Proof of proposition 4.3

The mixed algorithm of least squares and stochastic gradient learning can be associated to the big ode

$$
\begin{aligned}
\frac{d\Phi_A}{d\tau} &= R_A^{-1} M\left(g\left(\mathbf{\Phi}\right)\right)\left[T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_A\right] \\
\frac{dR_A}{d\tau} &= M\left(g\left(\mathbf{\Phi}\right)\right) - R_A \\
\frac{d\Phi_B}{d\tau} &= M\left(g\left(\mathbf{\Phi}\right)\right)\left[T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_B\right]
\end{aligned}
$$

The local stability of an REE $\Phi_f$ is therefore determined by the vectorised version of the small ode

$$
\frac{d\mathbf{\Phi}}{d\tau} = \left( \begin{array}{cc} \frac{d\Phi_A}{d\tau}, & \frac{d\Phi_B}{d\tau} \end{array} \right) = \left( \begin{array}{cc} T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_A, & M\left(g\left(\mathbf{\Phi}\right)\right)\left[T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_B\right] \end{array} \right)
$$

Therefore the relevant Jacobian is

$$
\begin{aligned}
J_3\left(\Phi_f\right) &= \frac{d}{d\ \mathrm{vec}\mathbf{\Phi}} \left( \begin{array}{c} \mathrm{vec}\ \left[T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_A\right] \\ \mathrm{vec}\ \left\{M\left(g\left(\mathbf{\Phi}\right)\right)\left[T\left(g\left(\mathbf{\Phi}\right)\right)' - \Phi_B\right]\right\} \end{array} \right)\Bigg|_{\mathbf{\Phi}=\left(\Phi_f,\Phi_f\right)} \\
&= \left( \begin{array}{cc} \frac{d\ \mathrm{vec}\left[T(g(\mathbf{\Phi}))'-\Phi_A\right]}{d\ \mathrm{vec}\Phi_A} & \frac{d\ \mathrm{vec}\left[T(g(\mathbf{\Phi}))'-\Phi_A\right]}{d\ \mathrm{vec}\Phi_B} \\ \frac{d\ \mathrm{vec}\left\{M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))'-\Phi_B\right]\right\}}{d\ \mathrm{vec}\Phi_A} & \frac{d\ \mathrm{vec}\left\{M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))'-\Phi_B\right]\right\}}{d\ \mathrm{vec}\Phi_B} \end{array} \right)\Bigg|_{\mathbf{\Phi}=\left(\Phi_f,\Phi_f\right)} \\
&= \left( \begin{array}{cc} \frac{d\ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d\ \mathrm{vec}\Phi_A} - I_{n_1 n_2} & \frac{d\ \mathrm{vec}T(g(\mathbf{\Phi}))'}{d\ \mathrm{vec}\Phi_B} \\ \frac{d\ \mathrm{vec}\left\{M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))'-\Phi_B\right]\right\}}{d\ \mathrm{vec}\Phi_A} & \frac{d\ \mathrm{vec}\left\{M(g(\mathbf{\Phi}))\left[T(g(\mathbf{\Phi}))'-\Phi_B\right]\right\}}{d\ \mathrm{vec}\Phi_B} \end{array} \right)\Bigg|_{\mathbf{\Phi}=\left(\Phi_f,\Phi_f\right)}
\end{aligned}
$$

Using similar arguments as in appendix A, it follows that

$$
\begin{aligned}
J_4(\Phi_f) &= \begin{pmatrix} \psi L(\Phi_f) - I_{n_1 n_2} & (1-\psi)L(\Phi_f) \\ \left[I_{n_1} \nwarrow M(\Phi_f)\right] \psi L(\Phi_f) & \left[I_{n_1} \nwarrow M(\Phi_f)\right]\left[(1-\psi)L(\Phi_f) - I_{n_1 n_2}\right] \end{pmatrix} \\
&= \begin{pmatrix} I_{n_1 n_2} & \mathbf{0} \\ \mathbf{0} & I_{n_1} \nwarrow M(\Phi_f) \end{pmatrix} \begin{pmatrix} \psi L(\Phi_f) - I_{n_1 n_2} & (1-\psi)L(\Phi_f) \\ \psi L(\Phi_f) & (1-\psi)L(\Phi_f) - I_{n_1 n_2} \end{pmatrix} \\
&= \begin{pmatrix} I_{n_1 n_2} & \mathbf{0} \\ \mathbf{0} & I_{n_1} \nwarrow M(\Phi_f) \end{pmatrix} J_1^{LS}(\Phi_f)
\end{aligned}
$$

■