Working Paper

On Market Dependencies of Agents' Learning for a Hyperinflation Model

Karl Henrik Johansson

WP-93-47
September 1993

IIASA
International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

On Market Dependencies of Agents' Learning for a Hyperinflation Model

Karl Henrik Johansson

WP-93-47
September 1993

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Abstract

This paper discusses an economic model for hyperinflation considered by Marcet and Sargent in [11]. The model describes the relation between the current price level, the money supply, and the agents' forecasting of the future price. The agents' learning is described by an ARMA-model which is fitted to the available series of old prices. It is shown that the agents' learning rate depends upon the inertia of the market, and an implicit formula is given for this dependence. A generalization of the hyperinflation model is also discussed.

Contents

On Market Dependencies of Agents' Learning for a Hyperinflation Model

*Karl Henrik Johansson**

1 Introduction

This paper deals with a model that shows the relation between changes in the money supply and the price level during hyperinflation. The model states back to Cagan [4], but has been further analyzed and extended *e. g.* by Fourgeaud *et al.* [5] and Marcet and Sargent [11]. Cagan's definition states that hyperinflation begins if prices rise more than 50 percent in a month. His model is developed from a study of seven hyperinflations occurring in different european countries during the period 1920–1946. He advances the hypothesis that during hyperinflations, the price level $P(k)$ at time k depends only on the expected price and the money supply $M(k)$.

The hypothesis of *rational expectations* is defined by Muth [12] to be that economic agents do not make systematic forecasting errors. In our case this would mean that agents' expectations of an economic variable coincide with its mathematical conditional expectation. However, in the model of this paper the *convergence* to rationality is studied as well. This requires the theory of stochastic approximation for the analysis, see *e. g.* [3].

We now briefly introduce the model set up of Marcet and Sargent [11]. Let us define $y(k) := \log P(k)$ and $x(k) := \log M(k)$. Further denote the agents' conditional expectation of $y(k+1)$ given information about the price level up to time k by $\hat{y}(k+1)$, *i. e.*

$$\hat{y}(k+1) := E\{y(k+1)|y(k), y(k-1), \ldots\}$$

The model of the market is now given by

$$y(k) = \lambda\hat{y}(k+1) + x(k) + v(k) \tag{1}$$

λ is a real constant parameter between zero and one (see [4] and [11]), and $\{v(k)\}$ is a white Gaussian stationary process with zero-mean and variance σ_v^2. v models the uncertainty in the relation between the quantities. The agents prediction of $y(k+1)$ is based upon an autoregressive model with external inputs (ARX-model). We assume that their learning about this model can be described by a recursive algorithm.

The parameter λ in (1) can be interpreted as the inertia of the market. This follows from the fact that both $\{x(k)\}$ and $\{v(k)\}$ are zero-mean stationary processes, so in average $y(k)$ will be equal to $\lambda \cdot \hat{y}(k+1)$.

Our main contribution in this paper will be to show that the agents' learning rate depends upon the inertia of the market. We show this by examining how λ influences the asymptotic variance of the agents' estimate. An implicit expression for this relation will be given. There will also be an attempt to generalize the model.

In Section 2 in this paper we will introduce a complete hyperinflation model including the corresponding learning algorithm. Section 3 treats a method for analyzing the type of economic models that we are studying. This method is used for deriving the agents' convergence point. The

*Department of Automatic Control, Lund Institute of Technology, Sweden. YSSP at IIASA 1993.

convergence rate of the learning algorithm is dealt with in Section 4, where also the connection between the market and the learning process is given. In Section 5 an introduction to a more general model is given, followed by the conclusions and some ideas for future work in Section 6. Appendix A gives the assumptions for the learning algorithm to have almost sure convergence, and in Appendix B is a new result about the convergence rate included.

2 The Hyperinflation Model

In this section we complete the hyperinflation model by describing how the money supply is modeled and how the agents' prediction is constructed. Thus, the equations for $\hat{y}$ and x in (1) are given. We will follow the structure of the model given in [11].

Let the money supply be modeled such that $x(k)$ satisfies the following autoregressive-moving average-equation (ARMA-equation)

$$x(k) + cx(k-1) = u(k) + du(k-1) \tag{2}$$

where $\{u(k)\}$ is a white Gaussian stationary process with zero-mean and variance σ_u^2. (Note that it is possible to include the noise term v from (1) in (2) without any lack of generality. We do not do that here, simply in order to compare our results with those in [11].)

The agents do not know the market model (1). They can only observe the price level or equivalently $\{y(k)\}$, and they think $y(k)$ is the output from a first order ARX-model

$$y(k) + ay(k-1) = w(k) + bw(k-1) \tag{3}$$

where $\{w(k)\}$ is believed by the agents to be an independent white Gaussian stationary process. The agents' prediction of $y(k+1)$ is now constructed by

$$\begin{aligned} \hat{y}(k+1) &= E\{y(k+1)|y(k), y(k-1), \ldots\} && (4) \\ &= -\hat{a}y(k) + \hat{b}\hat{w}(k) && (5) \end{aligned}$$

where $\hat{a}$ and $\hat{b}$ are the agents' estimate of a and b in (3). $\hat{w}(k)$ is given by filtering $\{y(k), y(k-1), \ldots\}$ through the filter given by (3). If we introduce the forward-shift operator q, which has the property $qf(k) = f(k+1)$, the filtering can be written as

$$\hat{w}(k) = \frac{q+\hat{a}}{q+\hat{b}} y(k) \tag{6}$$

where

$$\frac{A(q)}{B(q)} := \frac{q+\hat{a}}{q+\hat{b}}$$

is called the pulse transfer function, see *e.g.* [2].

The agents' learning of the unknown parameters a and b in (3) are approximated by a recursive learning algorithm. In our case we use a recursive pseudolinear regression (RPLR) algorithm (or extended least square algorithm), see [7]. For our convenience we collect the parameters and the regressors in the vectors

$$\theta(k) := \begin{bmatrix} -a(k) \\ b(k) \end{bmatrix} \qquad \varphi(k) := \begin{bmatrix} y(k-1) \\ \hat{w}(k-1) \end{bmatrix}$$

In the linear regression form, (3) will be

$$y(k) = \varphi^T(k)\theta + w(k) \tag{7}$$

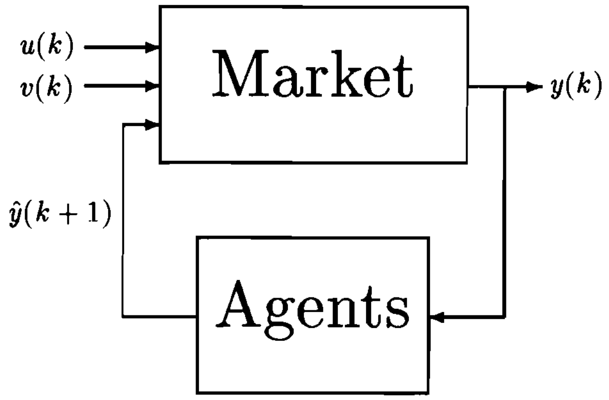

Figure 1: A block diagram that shows the relationship between the market and the agents.

If the estimate of θ is denoted by $\hat{\theta}$, the RPLR algorithm is given by the following three equations

$$\begin{aligned} \varepsilon(k) &= y(k) - \varphi^T(k)\hat{\theta}(k-1) \\ \hat{\theta}(k) &= \hat{\theta}(k-1) + \frac{1}{k}R^{-1}(k)\varphi(k)\varepsilon(k) \\ R(k) &= R(k-1) + \frac{1}{k}[\varphi(k)\varphi^T(k) - R(k-1)] \end{aligned} \tag{8}$$

where R asymptotically can be interpreted as the covariance matrix of $\hat{\theta}$.

To show almost sure convergence of the hyperinflation model, we need an algorithm that is slightly modified. (8) is changed so that if $(\hat{\theta}(k), R(k))$ happen to be outside a certain set when updated, they will instead be given values in the set. This new algorithm and the necessary assumptions for almost sure convergence are discussed in Appendix A.

In Figure 1 a block diagram shows how the market and the agents are connected. The block named "Market" represents the equations (1) and (2). Thus, the market is described by a linear time invariant system with three inputs ($u(k)$, $v(k)$, and $\hat{y}(k+1)$), and one output ($y(k)$). Because of the recursive learning algorithm running in the "Agents" block, the overall system is nonlinear and time variant. Notice that even if we are just considering the case when the agents have reached steady state (the parameters in the agents' model have converged), it is possible that the overall system for certain parameter values becomes unstable. This is due to the feedback of $y(k)$ through the agents' block, and is further discussed in Section 5.

3 The Limit State Convergence Point

We will now introduce a tool for analyzing economic models including learning algorithms. This will lead us to the limit state convergence point of the model. The formalism is developed by Sargent and Marcet, and is used in several of their papers (see *e.g.* [10], [11], [13], and [14]).

Define the state vector and the innovation vector

$$z(k) := \begin{bmatrix} y(k), & w(k), & x(k), & u(k) \end{bmatrix}^T \qquad e(k) := \begin{bmatrix} u(k), & v(k) \end{bmatrix}^T$$

Then we can collect (1), (2), and (3) in the state space equation

$$z(k+1) = T(\theta(k))z(k) + V(\theta(k))e(k) \tag{9}$$

Expressions for $T(\theta)$ and $V(\theta)$ are easily calculated and are given in [11]. We split $T(\theta)$:

$$T(\theta) = \begin{bmatrix} T_{11}(\theta) & T_{12}(\theta) \\ 0_{2\times 2} & T_{22}(\theta) \end{bmatrix}$$

where $T_{ij}(\cdot)$ are square matrices of dimension two, and $0_{2\times 2}$ is the zero matrix of dimension two. Further, let us for fixed θ introduce the covariance matrices

$$M(\theta) = \begin{bmatrix} M_{11}(\theta) & M_{12}(\theta) \\ M_{21}(\theta) & M_{22}(\theta) \end{bmatrix} := E\{z(k)z^T(k)\} \qquad \Omega := E\{e(k)e^T(k)\}$$

where M_{ij} are 2×2-matrices. M and Ω are symmetric by definition and M satisfies the Lyapunov equation

$$M(\theta) = T(\theta)M(\theta)T^T(\theta) + V(\theta)\Omega V^T(\theta)$$

(see [2]).

The agents assume that the true market model is given by the ARX-model (3). We will now compare the asymptotic estimates given this assumption, with the one given the true model (9). This is possible since we know that the learning algorithm asymptotically gives the orthogonal projection of $y(k+1)$ on $y(k)$ and $w(k)$. Denote the orthogonal projection of $\begin{bmatrix} y(k+1), & w(k+1) \end{bmatrix}^T$ on $\begin{bmatrix} y(k), & w(k) \end{bmatrix}^T$ by

$$S(\theta)\begin{bmatrix} y(k) \\ w(k) \end{bmatrix}$$

where $S(\cdot)$ is a 2×2-matrix with elements $S_{ij}(\cdot)$. $S(\cdot)$ is then the matrix that minimizes

$$E\left\{\left\| \begin{bmatrix} y(k+1) \\ w(k+1) \end{bmatrix} - S(\theta)\begin{bmatrix} y(k) \\ w(k) \end{bmatrix} \right\|\right\}$$

where $\|\cdot\|$ is the ordinary Euclidean norm. The minimum is given by (see [8])

$$S(\theta) = E\left\{ \begin{bmatrix} y(k+1) \\ w(k+1) \end{bmatrix} \begin{bmatrix} y(k), & w(k) \end{bmatrix} \right\} \left(E\left\{ \begin{bmatrix} y(k) \\ w(k) \end{bmatrix} \begin{bmatrix} y(k), & w(k) \end{bmatrix} \right\} \right)^{-1}$$

where we have assumed that

$$M_{11}(\theta) = E\left\{ \begin{bmatrix} y(k) \\ w(k) \end{bmatrix} \begin{bmatrix} y(k), & w(k) \end{bmatrix} \right\}$$

has full rank. Since

$$E\left\{ \begin{bmatrix} y(k+1) \\ w(k+1) \end{bmatrix} \begin{bmatrix} y(k), & w(k) \end{bmatrix} \right\} = \begin{bmatrix} T_{11}(\theta), & T_{12}(\theta) \end{bmatrix} \begin{bmatrix} M_{11}(\theta) \\ M_{21}(\theta) \end{bmatrix}$$

we finally get

$$S(\theta) = \begin{bmatrix} T_{11}(\theta), & T_{12}(\theta) \end{bmatrix} \begin{bmatrix} M_{11}(\theta) \\ M_{21}(\theta) \end{bmatrix} M_{11}^{-1}(\theta) = T_{11}(\theta) + T_{12}(\theta)M_{21}(\theta)M_{11}^{-1}(\theta)$$

Hence, asymptotically the agents' perceived projection of $y(k+1)$ on $y(k)$ and $w(k)$ will be given by θ, while the actual projection is given by $[S_{11}(\theta),\ S_{12}(\theta)]^T$. The fixed point θ_f of $S(\cdot)$ will under certain assumptions give the limit state convergence point of the hyperinflation model. These assumptions are given in Appendix A. Thus, it is possible to derive the convergence point simply by solving the equation

$$\theta = S(\theta)$$

In [11] it is shown that θ_f is unique.

4 The Convergence of the Learning Process

In this section we will concentrate on the the agents' learning process. The convergence of the RPLR algorithm is discussed. The differential equation associated with the algorithm is introduced and a theorem concerning the asymptotic distribution of the estimates is given. It is shown that this distribution depends upon the inertia of the market, the parameter λ in (1). Further, we will illustrate our result in two examples.

To analyze the behavior of a recursive learning algorithm, it is convenient to introduce the ***associated differential equation***. This approach was suggested by Ljung in [6], see also [7] and [1]. For sufficiently large k, the step size $\gamma(k) := 1/k$ in (8) will be small. Thus the correction in $\hat{\theta}(k)$ will be small. By assuming $\hat{\theta}(\cdot)$ vary slowly, we can approximate it over a small time interval by its averaged value $\bar{\theta}$. Then we have the approximated version of (7):

$$y(k) = \varphi^T(k, \bar{\theta})\bar{\theta}$$

We substitute R in (8) by its average $\bar{R}$. The updating rules are now

$$\begin{aligned} \hat{\theta}(k) &= \hat{\theta}(k-1) + \gamma(k)\bar{R}^{-1}f(\bar{\theta}) & (10) \\ R(k) &= R(k-1) + \gamma(k)[g(\bar{\theta}) - \bar{R}] & (11) \end{aligned}$$

where

$$\begin{aligned} f(\bar{\theta}) &= E\{\varphi(k)[y(k) - \varphi^T(k)\bar{\theta}]\} & (12) \\ g(\bar{\theta}) &= E\{\varphi(k)\varphi^T(k)\} & (13) \end{aligned}$$

(10) and (11) will act almost like the algorithm (8) in a neighborhood to $\bar{\theta}$ and $\bar{R}$ when k is large. With a change of time scale this new algorithm can be interpreted as a difference approximation of the differential equations

$$\begin{aligned} \frac{d\theta_D}{dt} &= R_D^{-1}(t)f(\theta_D(t)) & (14) \\ \frac{dR_D}{dt} &= g(\theta_D(t)) - R_D(t) & (15) \end{aligned}$$

The true estimates $\hat{\theta}$ and R will asymptotically follow the trajectories θ_D and R_D of these associated differential equations.

For a matrix A we introduce the notation $\bar{A}$ as the top row of A. Then, for our hyperinflation model we have

$$\begin{aligned} f(\theta) &= E\{\varphi(k)[y(k) - \varphi^T(k)\theta]\} \\ &= E\Big\{\varphi(k)\Big(\begin{bmatrix} \bar{T}_{11}(\theta) & \bar{T}_{12}(\theta) \end{bmatrix} z(k-1) + \bar{V}(\theta)e(k-1) - \varphi^T(k)\theta\Big)\Big\} \end{aligned}$$

and by using the symmetry of the covariance matrix $M(\cdot)$ and that $M_{11}(\cdot)$ is assumed to have full rank

$$\begin{aligned} f(\theta) &= \begin{bmatrix} M_{11}(\theta) & M_{12}(\theta) \end{bmatrix} \begin{bmatrix} \bar{T}_{11}(\theta) & \bar{T}_{12}(\theta) \end{bmatrix}^T - M_{11}(\theta)\theta \\ &= M_{11}(\theta)\Big(\begin{bmatrix} S_{11}(\theta) \\ S_{12}(\theta) \end{bmatrix} - \theta\Big) \end{aligned}$$

Further,

$$g(\theta) = E\{\varphi(k)\varphi^T(k)\} = M_{11}(\theta)$$

Hence, the associated differential equations for the learning algorithm of the hyperinflation model are

$$\frac{d\theta}{dt} = R^{-1}(t)M_{11}(\theta)(\bar{S}^T(\theta) - \theta) \tag{16}$$
$$\frac{dR}{dt} = M_{11}(\theta) - R(t) \tag{17}$$

To study the behavior of these nonlinear differential equations, it is possible to linearize them around a stationary point. For the differential equations (14) and (15) the linearized system is (see [1] p. 270)

$$\frac{d}{dt}\begin{bmatrix} \theta - \theta_f \\ R - R_f \end{bmatrix} = \begin{bmatrix} g^{-1}(\theta)\frac{\partial f}{\partial \theta} & 0 \\ * & -I_{\eta\times\eta} \end{bmatrix}_{\theta=\theta_f} \begin{bmatrix} \theta - \theta_f \\ R - R_f \end{bmatrix}$$

where (θ_f, R_f) is a stationary point, and $*$ denotes elements we are not interested in. The system matrix above has $\eta = \dim\theta$ eigenvalues of -1, and the rest determined by

$$g^{-1}(\theta)\frac{\partial f}{\partial \theta}\bigg|_{\theta=\theta_f}$$

Thus, for the hyperinflation model we have two eigenvalues -1 and the other two are equal to the eigenvalues of the matrix

$$\frac{\partial}{\partial \theta}(\bar{S}^T(\theta) - \theta) = \frac{\partial}{\partial \theta}\begin{bmatrix} S_{11}(\theta) & S_{12}(\theta) \end{bmatrix}^T - I \tag{18}$$

We will now show that the latter two eigenvalues are important for the convergence and the convergence rate of the hyperinflation model.

Consider the parameter estimation part of the learning algorithm (8). Introduce the function $H(\cdot,\cdot)$ as the updating of the estimate excluding the gain sequence $\{\gamma(k)\}$, *i. e.*

$$H(\theta(k-1), z(k)) := R^{-1}(k)\varphi(k)[y(k) - \varphi^T(k)\theta(k-1)]$$

Define

$$h(\theta) := E\{H(\theta(k-1), z(k))\} = R^{-1}f(\theta)$$

Then for the hyperinflation model

$$h(\theta,\lambda) = R^{-1}M_{11}(\theta)\left(\begin{bmatrix} S_{11}(\theta) \\ S_{12}(\theta) \end{bmatrix} - \theta\right) = \begin{bmatrix} S_{11}(\theta) \\ S_{12}(\theta) \end{bmatrix} - \theta \tag{19}$$

where the last equality follows from (17), since $R = M_{11}(\theta)$ in steady state. In (19) we have emphasized the dependencies of $h(\cdot,\cdot)$ upon the parameter λ in (1). Let us denote the covariance of $H(\cdot,\cdot)$ at the equilibrium point θ_f by $D(\cdot)$, *i. e.*

$$D(\lambda) := E\{H(\theta_f, z(k))H^T(\theta_f, z(k))\}$$

Since $\varphi(k)$ in

$$E\{\varphi(k)\varphi^T(k)w^2(k)\} \tag{20}$$

by definition just depends on $k-1, k-2, \ldots$, we can take the conditional expectation and split (20):

$$E\{\varphi(k)\varphi^T(k)w^2(k)\} = M_{11}(\theta_f)E\{w^2(k)\} = M_{11}(\theta_f)M_{11_{22}}(\theta_f)$$

where $M_{11_{22}}(\cdot)$ is the $(2,2)$-element in the matrix $M_{11}(\cdot)$. Hence

$$D(\lambda) = M_{11}^{-1}(\theta_f)M_{11_{22}}(\theta_f)$$

Introduce the notation $\lambda_i(A)$ for the ith eigenvalue of a matrix A and define (the matrix)

$$h_\theta := \frac{\partial h}{\partial \theta}$$

(Note the different meaning of λ and $\lambda_i(\cdot)$.) We now have the following theorem from [3].

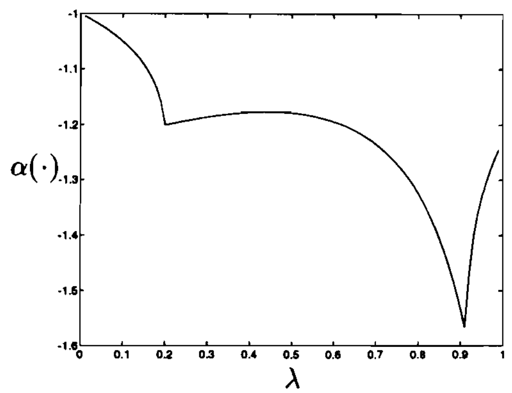

Figure 2: $\alpha(\cdot)$, the maximal real part eigenvalue of $h_\theta(\theta_f, \cdot)$, as a function of λ for Example 1. The cusps come from the eigenvalues shifting between being real and complex valued.

Theorem 1 *Assume the assumptions in Appendix A hold. If*

$$\alpha(\lambda) := \max_i Re\{\lambda_i(h_\theta(\theta_f, \lambda))\} < -1/2$$

then

$$\sqrt{k}(\theta(k) - \theta_f) \longrightarrow \mathcal{N}(0, P(\lambda)), \quad k \to \infty$$

where $P(\cdot)$ satisfies the Lyapunov equation

$$\left(\frac{1}{2}I + h_\theta(\theta_f, \lambda)\right)P(\lambda) + P(\lambda)\left(\frac{1}{2}I + h_\theta(\theta_f, \lambda)\right)^T = -D(\lambda)$$

□

Thus the eigenvalues of $h_\theta(\cdot,\cdot)$ are crucial for the behavior of the RPLR algorithm. In particular, it is enough to study the eigenvalue with the largest real part. Further, from (18) and (19) we notice that a necessary condition for the learning algorithm to be stable is that $\alpha(\lambda) < 0$. Hence, the assumption in Theorem 1 can be interpreted as that the learning algorithm has to be "sufficiently stable".

When $\alpha(\lambda) \in (-1/2, 0)$, Theorem 1 does not hold. A simulation method is then suggested in [11]. An assumption is made about what the convergence will look like. We show in Appendix B that it is possible in an analytical way to derive the convergence rate.

Theorem 1 gives an implicit formula for how the parameter λ in (1) influences the asymptotic convergence rate. We now use this formula in two example. This is an extension of [11]. We will consider two sets of parameters in the examples below, and show in some plots what the covariance matrix $P(\cdot)$ in Theorem 1 might look like. The notation Tr $P(\cdot)$ will be used for the trace of the matrix $P(\cdot)$, and Tr $P(\cdot)$ captures the essential feature of the corresponding Gaussian distribution.

Example 1 *Consider the hyperinflation model with parameters*

$$c = -0.9 \quad \sigma_v^2 = \sigma_u^2 = 0.01 \quad d = 0$$

These are the same parameters as in the example discussed on p. 27 in [11]. In Figure 2 it is shown how $\alpha(\cdot)$ (defined in Theorem 1) depends on λ. The two cusps are due to the eigenvalues shifting from real to complex and vice versa. If we further calculate $P(\cdot)$ through the formula in Theorem 1 for different values of λ and plot its trace as a function of λ, we get the result in Figure 3. Hence the asymptotic variance is a decreasing function of λ. An economic interpretation of this is that it is easier for the agents to learn if the inertia of the market is high. □

Let us treat another parameter set.

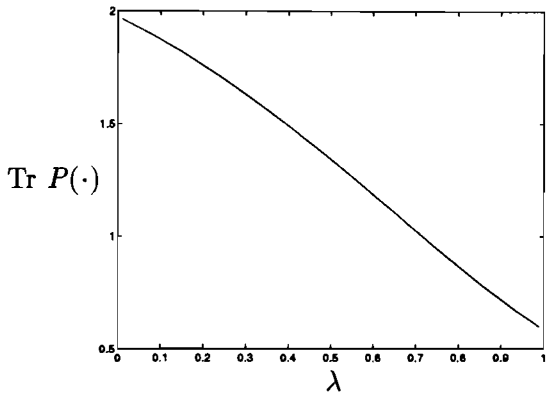

Figure 3: The trace of the asymptotic variance matrix $P(\cdot)$ shown as a function of λ for Example 1.

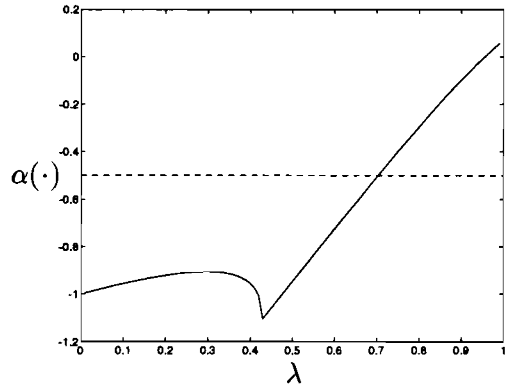

Figure 4: In Example 2, the maximal real part eigenvalue $\alpha(\cdot)$ exceeds $-1/2$ (the dashed line). Thus, for $\lambda > 0.7$ Theorem 1 is not useful.

Example 2 *On p. 21 in [11] simulations are given for the hyperinflation model with parameters*

$$c = -0.8 \quad \sigma_v^2 = \sigma_u^2 = 1 \quad d = -0.95$$

We treat this system in the same way as in previous example. The largest eigenvalue of $h_\theta(\theta_f, \cdot)$ is given in Figure 4. For $\lambda > 0.7$, we see that $\alpha(\lambda) > 1/2$ (the dashed line). Thus, Theorem 1 will not hold for $\lambda > 0.7$. (The cusp comes from the fact that the eigenvalues go from real to complex.) Tr $P(\cdot)$ is given in Figure 5 for $\lambda \in (0, 0.7)$. The variance is increasing tremendously when λ approaches 0.7. This example oppose the previous one, since it is now harder for the agents to learn if the market inertia is high. □

For most parameter choices above the estimate error decreases in time as $1/\sqrt{k}$. However, the limit Gaussian distribution has different covariance matrices which depend upon λ. In the second example the variance gets equal to infinity for $\lambda > 0.7$; the normalizing sequence $\sqrt{k}$ does not fit. Then we need the "slower" sequence k^γ, where $\gamma \in (0, -\alpha(\cdot))$ is arbitrary (see Appendix B).

Consequently, we have demonstrated that the inertia can have both positive and negative influence on the agents' learning rate. Note that the second example above contradicts the conjecture in [11] that $\alpha(\cdot)$ is always less than $-1/2$ for the hyperinflation model.

5 The Hyperinflation Model in Steady State

In this section we will discuss steady state behavior of the hyperinflation model given by (1). By this we mean that the parameters that the agents are learning have reached their stationary values. Further, it is assumed that these parameters are a fixed point of the corresponding function $S(\cdot)$ defined in Section 3. We will restate some of the results in the first part of [11]

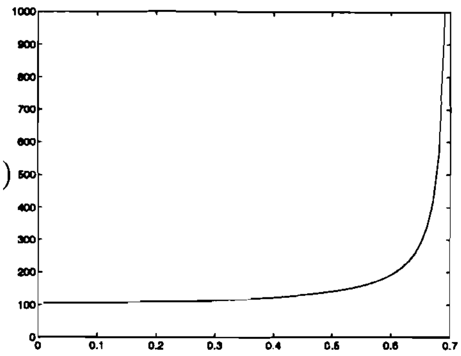

Figure 5: When λ approaches 0.7, Tr $P(\cdot)$ increases tremendously in Example 2. For $\lambda > 0.7$ the formula for calculating $P(\cdot)$ is not valid.

for a more general system. The generalization is that we allow the agents to use an ARX-model of arbitrary order, as well as we allow us to model the money supply x by an arbitrary ARMA(m,m)-model. This might let the hyperinflation model fit better to real data.

Consider (1) again. The money supply is modeled by the ARMA(m,m)-model

$$C(q)x(k) = D(q)u(k) \tag{21}$$

where

$$\begin{aligned} C(q) &:= q^m + c_1 q^{m-1} + \ldots + c_m & (22) \\ D(q) &:= q^m + d_1 q^{m-1} + \ldots + d_m & (23) \end{aligned}$$

C and D are assumed to be stable, *i. e.* all their zeros lie within the unit circle. Further, we assume that they have no common factors.

Assume that the agents use an ARX(n,n)-model to construct their prediction $\hat{y}(k+1)$. Thus they fit the difference equation

$$y(k+n) + a_1 y(k+n-1) + \ldots + a_n y(k) = w(k+n) + b_1 w(k+n-1) \ldots + b_n w(k) \tag{24}$$

to the data $\{y(k), y(k-1), \ldots\}$, or

$$A(q)y(k) = B(q)w(k) \tag{25}$$

where

$$\begin{aligned} A(q) &:= q^n + a_1 q^{n-1} + \ldots + a_n \\ B(q) &:= q^n + b_1 q^{n-1} + \ldots + b_n \end{aligned}$$

Both A and B are assumed to be stable, and that they have no common factors. Now the agents' prediction is given as

$$\begin{aligned} \hat{y}(k+n) &= E\{y(k+n)|y(k+n-1), y(k+n-2), \ldots\} \\ &= -a_1 y(k+n-1) - \ldots - a_n y(k) + b_1 \hat{w}(k+n-1) + \ldots + b_n \hat{w}(k) & (26) \\ &= (q^n - A(q))y(k) + (B(q) - q^n)\hat{w}(k) & (27) \end{aligned}$$

$\{\hat{w}(k)\}$ is by the agents believed to be the innovation in $y(k)$ relative to the information set $\{y(k-1), y(k-2), \ldots\}$. In (26) and (27) $\hat{w}(k)$ is the output of the filter given by (25) when $\{y(k), y(k-1), \ldots\}$ is the input, or if we write it in the same way as in (6),

$$\hat{w}(k) = \frac{A(q)}{B(q)} y(k)$$

Recall the function $S(\cdot)$ from Section 3. We use it in a definition from [11]:

Definition 1 *A stationary limited information rational expectations equilibrium (LIREE) is a fixed point of* $S(\cdot)$ *that generates invertible stable pulse transfer functions from* $(\{x(k), v(k)\})$ *to* $(\{y(k)\}, \{w(k)\})$ *and from* $\{y(k)\}$ *to* $\{w(k)\}$. □

By inserting (26) and (21) in (1), we get

$$(B(q) - \lambda q B(q) + \lambda q A(q))C(q)y(k) = B(q)D(q)u(k) \tag{28}$$

and thus the following generalization of Proposition 2 in [11]:

Lemma 1 *Assume that*

$$\theta = \begin{bmatrix} -a_1 & \dots & -a_n & b_1 & \dots & b_n \end{bmatrix}^T \tag{29}$$

is a fixed point of $S(\cdot)$ *and that there are no common factors between (30) and* B *or* D. θ *will be a LIREE if and only if*

$$B(q) - \lambda q B(q) + \lambda q A(q) \tag{30}$$

is stable.

We illustrate the lemma by considering the model in the earlier sections.

Example 3 *Let the money supply being modeled by an ARMA(1,1)-model and let the agents use an ARX(1,1)-model for the prediction. Thus*

$$\begin{aligned} A(q) &= q + a & B(q) &= q + b \\ C(q) &= q + c & D(q) &= q + d \end{aligned}$$

Now Lemma 1 gives that

$$(1 - \lambda b + \lambda a)q + b$$

must be stable or, equivalent with [11], it must hold that

$$\left| \frac{b}{1 - \lambda b + \lambda a} \right| < 1$$

□

Note that the analysis here is built upon the fact that the recursive algorithm of the agents converges to the correct values. This will happen in some cases, for example for the system in the example above. It is an interesting question to find out for which polynomials A, B, C, and D we have convergence.

6 Conclusions and Ideas for Future Research

We have been studying a model for hyperinflation. This model describes the interaction between the market and the agents. In particular, we have discussed how the learning of the agents depends upon the market. In a theorem it was shown how, in an implicit way, the convergence rate of the agents learning algorithm is related to the inertia of the market. Further, two examples illustrated that it can be hard for the agents to learn about the market both if the market inertia is high as well as if it is low.

An attempt to generalize the model has also been done. In particular, we considered ARMA-models for the money supply description of an arbitrary order. The agents were allowed to assume a market model of any ARX(n,n)-type.

Finally, we will mention some ideas for future work. It is desirable to get a better understanding of the observed relations between the agents' learning rate and the market inertia; in economic terms as well as in mathematical. Another interesting topic is to try to fit or compare the model to real data. We also believe that it is possible to proof many of the convergence results in [11] for the more general model in Section 5.

7 Acknowledgments

I want to express my gratitude to my supervisor at IIASA Dr. Yuri M. Kaniovski. He has encouraged me throughout my work.

A Assumptions

In this appendix we will give the assumptions for having almost sure convergence of the learning algorithm in Section 2. These are also the assumptions for Theorem 1 in Section 4. We just repeat the assumptions given in [11].

Define the sets

$$D_S := \{\theta \mid |\lambda_i(T(\theta))| < 1 \ \forall i\}$$

and D_A as the domain of attraction of a fixed point of the differential equations (16) and (17). Also, define the open and bounded set D_1 and the closed set D_2 by the relations

$$D_2 \subset D_1 \subset \mathbb{R}^{2\times 2^2}$$

and

$$(\theta, R) \in D_1 \quad \Rightarrow \quad \theta \in D_S$$

We consider the following modification of the RPLR-algorithm (8)

$$\begin{aligned} \theta_1(k) &= \theta(k-1) + \frac{1}{k}R^{-1}(k)\varphi(k)[y(k) - \varphi^T(k)\theta(k-1)] \\ R_1(k) &= R(k-1) + \frac{1}{k}[\varphi(k)\varphi^T(k) - R(k-1)] \\ (\theta(k), R(k)) &= \begin{cases} (\theta_1(k), R_1(k)) & (\theta_1(k), R_1(k)) \in D_2 \\ (\theta(k-1), R(k-1)) & (\theta_1(k), R_1(k)) \notin D_2 \end{cases} \end{aligned} \tag{31}$$

which assures us that the estimates always stay in the set D_1. The following assumptions will give almost sure convergence for $\{\theta(k)\}$.

1. $S(\cdot)$ has a unique fixed point in D_S.
2. Each element in $T(\theta)$ is two times differentiable and each element in $V(\theta)$ is one time differentiable for all $\theta \in D_S$.
3. $M_{11}(\cdot)$ defined in Section 3 has full rank.
4. For $\{e(k)\}$ in (9) it is true that $E\{|e(k)|^p\}$ for all $p > 1$.
5. There exists a subset Ω_0 of the sample space such that $\Pr\{\Omega_0\} = 1$. There also exists two random variables $\bar{C}_1(\omega)$ and $\bar{C}_2(\omega)$ and a subsequence $\{k_h\}$ such that

$$\begin{aligned} |z(k_h)| &< \bar{C}_1(\omega) \\ |R(k_h)| &< \bar{C}_2(\omega) \end{aligned}$$

   for all $\omega \in \Omega_0$ and $h = 1, 2, \ldots$.
6. The trajectories of the associated differential equations (16) and (17) with initial conditions $(\theta(0), R(0)) \in D_2$ do not leave D_1.

B An Analytical Result for the Convergence Rate

We will now give an an estimate for the convergence rate of the recursive learning algorithm when the algorithm is not "stable enough", *i. e.* that the eigenvalues assumption in Theorem 1 does not hold. Sargent and Marcet *assume* what the convergence rate is in the case Theorem 1 does not hold. We will in this appendix derive analytical results, and show that their conjecture is right. To some extension, we will refer to the assumptions and proof of Theorem 24 (p. 246) in [3].

Introduction

Consider the recursive algorithm (8) again

$$\begin{aligned}\theta(k) &= \theta(k-1) + \frac{1}{k}H(\theta(k-1), z(k)) \\ &= \theta(k-1) + \frac{1}{k}h(\theta(k-1)) + \frac{1}{k}\epsilon(k)\end{aligned} \tag{32}$$

where $z(\cdot)$ is the state in the state space description (9), and the expression for $\epsilon(k)$ is given in [3].

Given the assumptions and the modification of the algorithm in Appendix A, we know that $\{\theta(k)\}$ converges to θ_f almost surely. Further, if the assumption of Theorem 1 in Section 4 holds, we know that

$$\sqrt{k}(\theta(k) - \theta_f) \longrightarrow \mathcal{N}(0, P), \quad k \to \infty$$

We will show in this appendix that it is possible to determine an estimate of the convergence rate even if the eigenvalues assumption in Theorem 1 does not hold. Denote (as in Section 4)

$$\alpha := \max_i \mathrm{Re}\ \{\lambda_i(h_\theta(\theta_f))\}$$

If $\alpha \in (-1/2, 0)$, we will show that

$$k^\gamma(\theta(k) - \theta_f) \longrightarrow \mathcal{D}, \quad k \to \infty$$

where γ is arbitrary such that $\gamma < |\alpha|$. $\mathcal{D}$ is the distribution which puts unit mass at zero.

Analysis

We will need the definitions for almost sure convergence, convergence in probability, and convergence in quadratic mean. Thus, recall:

Definition 2 *$\{\theta(k)\}$ converges almost surely to θ_f if $\forall\varepsilon > 0$ and $\forall\delta > 0$, $\exists N(\varepsilon, \delta)$:*

$$\Pr\{\|\theta(k) - \theta_f\| > \varepsilon, k \geq m\} < \delta, \quad \forall m \geq N(\varepsilon, \delta) \tag{33}$$

where $\|\cdot\|$ is the Euclidean norm. We denote this

$$\theta(k) \xrightarrow{a.s.} \theta_f$$

Definition 3 *$\{\theta(k)\}$ converges in probability to θ_f if $\forall\varepsilon > 0$ and $\forall\delta > 0$, $\exists N(\varepsilon, \delta)$ such that if $k \geq N(\varepsilon, \delta)$*

$$\Pr\{\|\theta(k) - \theta_f\| > \varepsilon\} < \delta$$

which is denoted

$$\theta(k) \xrightarrow{p.} \theta_f$$

□

Definition 4 *$\{\theta(k)\}$ converges in quadratic mean to θ_f if*

$$E\{(\theta(k)-\theta_f)^2\} \to 0, \quad k \to \infty$$

We denote this

$$\theta(k) \xrightarrow{q.m.} \theta_f$$

□

Introduce the function

$$\tilde{h}(\theta) := \begin{cases} h(\theta), & \|\theta-\theta_f\| \leq \varepsilon \\ h_\theta(\theta_f)(\theta-\theta_f), & \|\theta-\theta_f\| > \varepsilon \end{cases}$$

and the new algorithm

$$\begin{aligned} \tilde{\theta}(k) &= \tilde{\theta}(k-1) + \frac{1}{k}\tilde{H}(\theta(k-1), z(k)) \\ &= \tilde{\theta}(k-1) + \frac{1}{k}\tilde{h}(\theta(k-1)) + \frac{1}{k}\tilde{\epsilon}(k) \end{aligned} \tag{34}$$

where $\tilde{H}(\cdot,\cdot)$ and $\tilde{\epsilon}$ can be calculated from (32). It is easily checked that $\{\tilde{\theta}(k)\}$ converges almost surely to $\{\theta_f\}$. It follows that if $\tilde{\theta}(N(\varepsilon,\delta)) = \theta(N(\varepsilon,\delta))$, then for $k \geq N(\varepsilon,\delta)$ the algorithm (34) will act like (32) with probability at least $1-\delta$. Convergence in quadratic mean will be shown for the algorithm (34), and that this implies convergence in probability for the original learning algorithm (32) modified as in Appendix A.

By showing how the conditional variance of the estimates in the algorithm (34) progress, we are able to tell the behavior of the algorithm. Let us introduce the σ-field generated by $z(1), z(2), \ldots, z(k)$ and denote it F_k. Define

$$\Delta(k) := E\{(\tilde{\theta}(k)-\theta_f)^2 \mid F_{k-1}\}$$

We call a matrix A stable if it has all real parts of its eigenvalues less than zero. Recall the well-known Lyapunov theorem

Lemma 2 *Assume A is a stable matrix. Then, for every positive definite matrix Q, there exist a symmetric positive definite matrix P such that*

$$A^T P + PA = -Q$$

□

Given two vectors x and y introduce the inner product

$$\langle x, y\rangle := x^T y$$

Define a second inner product out of the first and the matrix P in Lemma 2:

$$[x, y] := \langle Px, y\rangle \tag{35}$$

Then we have the following lemma:

Lemma 3 *Assume the matrices in Lemma 2 exists. Then for all x*

$$[Ax, x] < 0$$

□

Proof: From Lemma 2 we have

$$\langle A^T Px, x\rangle + \langle PAx, x\rangle = -\langle Qx, x\rangle$$

Since

$$\langle A^T Px, x\rangle = \langle Px, Ax\rangle = [x, Ax] = [Ax, x]$$

we conclude that

$$2[Ax, x] = -\langle Qx, x\rangle < 0$$

□

We know that $\tilde{h}_\theta(\theta_f)$ is a stable matrix. Thus, $\tilde{h}_\theta(\theta_f)+\gamma I$ is also stable for all $\gamma < -\alpha$. Lemma 3 gives the following upper limit

$$[\tilde{h}_\theta(\theta_f)x, x] < -\gamma[x, x]$$

Let us repeat the introductory part of the proof of Theorem 24 (p. 256) in [3] for our learning algorithm. The norm used is the one defined by the inner product $[\cdot,\cdot]$. We can state an inequality similar to (1.10.16) in [3]:

$$\Delta(k) \leq (1 - \frac{2\gamma}{k})\Delta(k-1) + \frac{C_1}{k^2} + \frac{2}{k}E\{f(k-1) - f(k)\} \tag{36}$$

Throughout this appendix C_i denotes constants. For the stochastic process $\{f(k)\}$, it is possible to show that

$$E\{|f(k)|\} \leq C_2 \tag{37}$$

(see [3]). By iterating (36), we get

$$\begin{aligned} \Delta(k) \;&\leq\; \Delta(0)\prod_{i=1}^{k}(1-\frac{2\gamma}{i}) + C_1\sum_{i=1}^{k}\frac{1}{i^2}\prod_{j=i+1}^{k}(1-\frac{2\gamma}{j}) \\ &\quad +2\sum_{i=1}^{k}\frac{1}{i}E\{f(i-1)-f(i)\}\prod_{j=i+1}^{k}(1-\frac{2\gamma}{j}) \\ &=: \; T_1 + T_2 + T_3 \end{aligned}$$

Above as well as below, we follow the convention that

$$\prod_{j=i}^{k}(\cdot) = 1$$

when $i > k$. We treat the three terms T_1, T_2, and T_3 separately. Before we show the convergence of these terms, we recall some inequalities. For large k we have

$$\prod_{i=1}^{k}(1-\frac{2\gamma}{i}) \leq \exp\Big(-2\gamma\sum_{i=1}^{k}\frac{1}{i}\Big) \leq \exp(-2\gamma\ln k) = k^{-2\gamma} \tag{38}$$

which gives

$$\begin{aligned} \sum_{i=1}^{k}\frac{1}{i^2}\prod_{j=i+1}^{k}(1-\frac{2\gamma}{j}) \;&=\; \sum_{i=1}^{k}\frac{1}{i^2}\prod_{j=1}^{k-i}(1-\frac{2\gamma}{j+i}) \leq \sum_{i=1}^{k}\frac{1}{i^2}\prod_{j=1}^{k-i}(1-\frac{2\gamma}{j}) \\ &\leq\; \sum_{i=1}^{k}\frac{1}{i^2}\frac{k^{-2\gamma}}{i^{-2\gamma}} = k^{-2\gamma}\sum_{i=1}^{k}i^{2\gamma-2} \end{aligned} \tag{39}$$

and for $\beta < 0$

$$\sum_{i=1}^{k} i^\beta \leq \int_0^k i^\beta di = \frac{k^{\beta+1}}{\beta+1} \tag{40}$$

Using (38), an estimation for T_1 is obtained:

$$T_1 \leq \Delta(0)k^{-2\gamma}$$

Since $\gamma \in (0, 1/2)$, we have $\beta := 2\gamma - 2 < 0$. Thus, (39) and (40) gives

$$T_2 \leq C_3 k^{-1}$$

T_3 needs more detailed investigation. Firstly, note that

$$\begin{aligned}
&\sum_{i=1}^{k} \frac{1}{i}[f(i-1) - f(i)] \prod_{j=i+1}^{k} (1 - \frac{2\gamma}{j}) \\
&= \sum_{i=1}^{k} \frac{1}{i}[f(i-1) - f(i)](1 - \frac{2\gamma}{i+1}) \prod_{j=i+2}^{k} (1 - \frac{2\gamma}{j}) \\
&= [f(0) - f(1)](1-\gamma)\prod_{j=3}^{k}(1 - \frac{2\gamma}{j}) - \sum_{i=2}^{k} \frac{f(i-1)}{i(i-1)} \prod_{j=i+2}^{k} (1 - \frac{2\gamma}{j}) \\
&\quad + \sum_{i=2}^{k} \left[\frac{f(i-1)}{i-1} - \frac{f(i)}{i}\right] \prod_{j=i+2}^{k} (1 - \frac{2\gamma}{j}) \\
&\quad - 2\gamma \sum_{i=1}^{k} \left[\frac{f(i-1)}{i(i+1)} - \frac{f(i)}{i(i+1)}\right] \prod_{j=i+2}^{k} (1 - \frac{2\gamma}{j}) \\
&=: S_1 + S_2 + S_3 + S_4
\end{aligned}$$

S_1 can be treated like T_1:

$$|S_1| \leq C_4 \prod_{j+3}^{k}(1 - \frac{2\gamma}{j}) \leq C_5 k^{-2\gamma}$$

We have by using (37)

$$|S_2| = \left|\sum_{i=2}^{k} \frac{f(i-1)}{i(i-1)} \prod_{j=i+2}^{k} (1 - \frac{2\gamma}{j})\right| \leq \sum_{i=1}^{k} \frac{C_6}{i^2} \prod_{j=i+1}^{k} (1 - \frac{2\gamma}{j}) \leq C_7 k^{-1}$$

Further

$$\begin{aligned}
S_3 &= \sum_{i=2}^{k} \frac{f(i-1)}{i-1} \prod_{j=i+2}^{k} (1 - \frac{2\gamma}{j}) - \sum_{i=2}^{k} \frac{f(i)}{i} \prod_{j=i+2}^{k} (1 - \frac{2\gamma}{j}) \\
&= f(1)\prod_{j=4}^{k}(1 - \frac{2\gamma}{j}) + \sum_{i=2}^{k} \frac{f(i)}{i} \prod_{j=i+3}^{k} (1 - \frac{2\gamma}{j}) - \sum_{i=2}^{k} \frac{f(i)}{i}(1 - \frac{2\gamma}{i+2}) \prod_{j=i+3}^{k} (1 - \frac{2\gamma}{j})
\end{aligned}$$

Thus

$$|S_3| \leq |f(1)|\prod_{j=4}^{k}(1 - \frac{2\gamma}{j}) + C_8 \sum_{i=2}^{k} \frac{1}{i^2} \prod_{j=i+3}^{k} (1 - \frac{2\gamma}{j}) \leq C_9 k^{-2\gamma} + C_{10} k^{-1}$$

S_4 is treated in the same way as S_2:

$$\begin{aligned}
|S_4| &\leq \left|\sum_{i=1}^{k} \frac{f(i-1)}{i(i+1)} \prod_{j=i+2}^{k} (1 - \frac{2\gamma}{j})\right| + \left|\sum_{i=1}^{k} \frac{f(i)}{i(i+1)} \prod_{j=i+2}^{k} (1 - \frac{2\gamma}{j})\right| \\
&\leq 2C_2 \sum_{i=1}^{k} \frac{1}{i^2} \prod_{j=i+2}^{k} (1 - \frac{2\gamma}{j}) \leq C_{11} k^{-1}
\end{aligned}$$

This ends the investigation of T_3.

To conclude, we have shown that

$$\Delta(k) \leq C_{12}k^{-2\gamma} + C_{13}k^{-1} \tag{41}$$

For large k, the first term in the right hand side above dominates. Thus

$$\lim_{k\to\infty} \sup k^{2\gamma}\Delta(k) \leq C_{12}$$

We will now finally show that the convergence in quadratic mean of $\{\tilde{\theta}(k)\}$ above implies convergence in probability of $\{\theta(k)\}$, *i. e.* convergence in probability for the original algorithm. We have shown that

$$k^{\gamma}(\tilde{\theta}(k) - \theta_f) \xrightarrow{q.m.} 0, \quad k \to \infty$$

This implies that

$$k^{\gamma}(\tilde{\theta}(k) - \theta_f) \xrightarrow{p.} 0, \quad k \to \infty$$

or equivalently that $\forall \sigma > 0$ and $\forall \mu > 0$, $\exists M(\sigma, \mu)$ such that

$$\Pr\{\|k^{\gamma}(\tilde{\theta}(k) - \theta_f\| > \sigma\} < \mu/2, \quad k \geq M(\sigma,\mu)$$

For $k \geq \max\{N(\varepsilon,\delta), M(\sigma,\mu)\}$ we have

$$\begin{aligned}
\Pr\{\|k^{\gamma}(\theta(k) - \theta_f)\| > \sigma\} &= \Pr\{\|k^{\gamma}(\theta(k) - \theta_f)\| > \sigma,\ \theta(k) = \tilde{\theta}(k)\} \\
&\quad + \Pr\{\|k^{\gamma}(\theta(k) - \theta_f)\| > \sigma,\ \theta(k) \neq \tilde{\theta}(k)\} \\
&\leq \Pr\{\|k^{\gamma}(\theta(k) - \theta_f)\| > \sigma,\ \theta(k) = \tilde{\theta}(k)\} + \Pr\{\theta(k) \neq \tilde{\theta}(k)\} \\
&\leq \Pr\{\|k^{\gamma}(\tilde{\theta}(k) - \theta_f)\| > \sigma\} + \delta = \mu/2 + \delta \leq \mu
\end{aligned}$$

where the last inequality is given if we in Equation (33) choose $\delta \leq \mu/2$. Hence, we have shown that

$$k^{\gamma}(\theta(k) - \theta_f) \xrightarrow{p.} 0, \quad k \to \infty$$

where $\gamma < -\alpha$.

It should be noted that the estimate given here is not the best possible, not even with respect to the order (in k) with which $\{\theta(k)\}$ converges to θ_f. For a discussion of this in the simplest Markov stochastic approximation case see [9]. Further, note that the inner product introduced in (35) is a function of λ and that we have not analyzed here how this new topology depends upon λ.

References

[1] K J Åström and B Wittenmark. *Adaptive control.* Addison-Wesley Publishing Company, 1989.

[2] K J Åström and B Wittenmark. *Computer-controlled systems—theory and design.* Prentice-Hall International, Inc., 2nd edition, 1990.

[3] A Benveniste, M Metivier, and P Priouret. *Adaptive algorithms and stochastic approximations.* Springer-Verlag, Berlin, Heidelberg, 1990.

[4] P Cagan. The monetary dynamics of hyperinflation. In M Freidman, editor, *Studies in the quantity theory of money*, pages 25–117. The University of Chicago Press, 1956.

[5] C Fourgeaud, C Gourieroux, and J Pradel. Learning procedures and convergence to rationality. *Econometrica*, 54(4), 1986.

[6] L Ljung. Analysis of recursive stochastic algorithms. *Transactions on Automatic Control*, 22(4), 1977.

[7] L Ljung and T Söderström. *Theory and practice of recursive identification.* M.I.T. Press, Cambridge, MA, 1983.

[8] D G Luenberger. *Optimization by vector space methods.* John Wiley & Sons, Inc., 1969.

[9] P Major and P Revesz. A limit theorem for the Robbins-Monro approximation. *Z. Wahrscheinlichkeitstheorie. verw. Geb.*, 27:79–86, 1973.

[10] A Marcet and T J Sargent. Convergence of least-squares learning in environments with hidden state variables and private information. *Journal of Political Economy*, 97(6):1306–1322, 1989.

[11] A Marcet and T J Sargent. Speed of convergence of recursive least squares learning with ARMA perceptions. 1992. Mimeographed.

[12] J F Muth. Rational expectations and the theory of price movements. *Econometrica*, 29(3), 1961.

[13] T J Sargent. Equilibrium with signal extraction from endogenous variables. *Journal of Economic Dynamics and Control*, 15, 1991.

[14] T J Sargent and A Marcet. Convergence of least squares learning mechanisms in self-referential linear stochastic models. *Journal of Economic Theory*, 48(2):337–368, Aug 1989.