Linear Difference Equations and Autoregressive Processes

© Marc-Andreas Muendler
University of California at Berkeley

February 17, 2000

1 Introduction

In this summary on difference equations and autoregressive processes you will find that many of the seemingly complicated mathematical entities, such as eigenvalues, diagonalized matrices, unit roots and non-stationary processes, are nothing special after all. I will show in sections 3 and 4 that eigenvalues are merely a little more advanced versions of the coefficient a in a linear difference equation $x_t = ax_{t-1}$. As is quite obvious for this difference equation already, x_t will grow bigger and bigger (explode) if a is larger than one. Similarly, eigenvalues that exceed unity make a system of difference equations explosive. As I will argue in section 5 for scalar difference equations, these eigenvalues are nothing but solutions to polynomials. At least for the case of a second-order difference equation, where a second-order polynomial is involved, we all know how to handle this. A second-order polynomial is a quadratic equation for which we used to find the solutions as early as in high school. Finally, I take these concepts to a stochastic setting and consider autoregressive processes in section 6. Again, it all reduces to very basic and known concepts. An autoregressive process is non-stationary exactly in the cases where a difference equation would be explosive. That is, an autoregressive process $x_t = ax_{t-1} + \tilde{\epsilon}_t$ is explosive when the little coefficient a exceeds one.

2 Linear first-order difference equations in one variable

When only one variable is concerned, a linear first-order difference equation takes the form

$$x_t = ax_{t-1} + b_t. \tag{1}$$

It is called a first-order difference equation because only one lag of x appears. In this equation, a is a time-independent coefficient and b_t is called the 'forcing term.' When $b_t = 0$, the simplest case, we call the difference equation homogeneous, and otherwise non-homogeneous. When $b_t = b$ for all t, the difference equation is said to be non-homogeneous but autonomous. Finally, when b_t is time-dependent, we call the difference equation non-homogeneous and non-autonomous. A solution to the system is a function of the coefficients and time that satisfies (1) (along with a so-called 'boundary condition'). As we will see soon, the general solution to the non-homogeneous, non-autonomous system is

$$\boxed{x_t = c \cdot a^t + \sum_{s=-\infty}^{t} a^{t-s} b_s \qquad \text{when } |a| < 1,} \tag{2}$$

and

$$\boxed{x_t = c \cdot a^t + \sum_{s=t}^{\infty} (\tfrac{1}{a})^{s-t} b_s \qquad \text{when } |a| > 1.} \tag{3}$$

The constant c is a real number to be determined. Note that the value of x_t is converging to $\sum_{s=-\infty}^{t} a^{t-s} b_s$ if $|a| < 1$. It is exploding otherwise. However, let's defer a discussion of stability until section 2.2 and first derive the solution.

2.1 Solving the difference equation

The single most useful principle to solve difference equations is called the *Superposition Principle*. It says, the general solution to any linear difference equation can be split up into to parts: First, the solutions to the homogeneous part of it and, second, *any* particular solution to the non-homogeneous difference equation you like to use:

$$\boxed{x_t^{general} = x_t^{complementary} + x_t^{particular}.} \tag{4}$$

The term $x_t^{complementary}$ denotes the solutions to the homogeneous part of the difference equation. As surprising as this may seem, you will soon be convinced that it must hold true for our equation (1). It is not much harder to prove for the general case.

Finding the solution to the homogeneous part of (1) is simple. The homogeneous part is the remainder of (1) when you set $b_t = 0$. For the solution of it, simply plug (1) into itself recursively. That is, write

$$x_t^{complementary} = ax_{t-1} = a(ax_{t-1}) = a(a(ax_{t-2})) = \ldots = a^t x_0. \qquad (5)$$

[You can check that this is a solution to (1) by plugging $a^{t-1}x_0$ back into (1) for x_{t-1} when $b_t = 0$.] The function $x_t^{complementary} = a^t x_0$ is a solution to (1) for *any* value of x_0 that you like to choose. So, $x_t = a^t x_0$ is in fact a whole family of solutions. We get to know all family members by varying through x_0 from negative to positive infinity. In the solutions (2) and (3), the x_0 has been replaced by a small c to indicate the arbitraryness of this constant. We can only pin down the c to a concrete value if we have a boundary (or initial condition) such as $x_0 = 1.7$ or $x_{-2000} = 42$.

Let's proceed to the solution of the non-homogeneous system. By the Superposition Principle, it suffices to find one single or *particular* solution to the non-homogeneous system, and we are done. The most convenient tool around for this are lag and lead operators. You just need to know when to apply which one. The rule is

> Whenever $|a| < 1$, keep the difference equation in form (1) and apply lag operators.
>
> Whenever $|a| > 1$, divide the difference equation (1) by a and apply lead operators.

The result will be one particular (but not the general solution) to the linear difference equation.

Let's take the case of $|a| < 1$. Rewrite the difference equation (1) in terms of lag operators,

$$x_t = a\mathbb{L}x_t + b_t,$$

or

$$(1 - a\mathbb{L})x_t = b_t.$$

Hence, by the rules of the lag-operator,

$$x_t^{particular} = \frac{b_t}{1 - a\mathbb{L}} = (1 + a\mathbb{L} + a^2\mathbb{L}^2 + a^3\mathbb{L}^3 + \dots)b_t = \sum_{s=-\infty}^{t} a^{t-s} b_s. \quad (6)$$

This is one particular solution to the non-autonomous linear difference equation (1). [You can verify that it is one by plugging $x_{t-1} = \sum_{s=-\infty}^{t-1} a^{t-1-s} b_s$ back into (1).] But it is not the only solution! From the Superposition Principle we know that we need to add the complementary solution in order to make things general. It is a surprising, but important property of the lag (and lead) operator that it eliminates the complementary term. Applying the Superposition Principle, and adding (6) to (5), we find the general solution, as it was reported in the very beginning in (2). [Check by plugging $x_{t-1} = c \cdot a^{t-1} + \sum_{s=-\infty}^{t-1} a^{t-1-s} b_s$ into (1)]. Similarly, if $|a| > 1$, following the same steps as above but using lead operators yields (3).

2.2 Stability

We say a difference equation is *stable* whenever the parameter a is strictly less than one in absolute value, that is whenever $|a| < 1$. It becomes clear from the solution to the homogeneous difference equation, (5), why we would call it stable in that case. Whenever $|a| < 1$, our x_t must decrease over time and converge to zero. Consider

$$x_t^{compl.} = a^t x_0$$

and suppose we start out with some $x_0 \neq 0$. Since $|a| < 1$, raising a to some power t makes the coefficient a^t smaller and smaller as t increases. This is true irrespective of whether a is positive or negative. Therefore $x_t^{compl.}$ converges to zero. It does so in a cycling pattern when a is negative since then the coefficient a^t is positive in every second period, and negative in every other period. When a is positive but less than one, $x_t^{compl.}$ converges straight to zero. So far so good for a homogeneous difference equation.

In the case of a non-homogeneous but autonomous difference equation (where $b_t = b \neq 0 \; \forall t$), stability would still only depend on the absolute value of a. For $|a| < 1$, our x_t would now converge to the so-called 'steady state' value. The 'steady state' is the value of x_t, usually denoted by $\bar{x}$, for which $x_t = \bar{x}$ forever. That is, the 'steady state' is the one state of our dynamic

system that will never be left if we reach it once. It is easy to find the steady state in our autonomous first-order difference equation $x_t = ax_{t-1} + b$. Setting $x_t = x_{t-1} = \bar{x}$, and solving out for $\bar{x}$, we find $\bar{x} = b/(1-a)$. As can also be seen from (2), the general solution in the homogeneous case is then $x_t^{general} = x_0 a^t + b/(1-a)$. [Check by simplifying $\sum_{s=-\infty}^{t} a^{t-s} b$ in (2).] Thus, whenever $|a| < 1$, the dynamic forces of the system take our x_t back to the steady-state no matter at what x_0 we started out.

Even in the case of a non-autonomous difference equation, stability mainly depends on the absolute value of a. If the sum $\sum_{s=-\infty}^{t} a^{t-s} b_s$ is finite, the power of the fact that $|a| < 1$ will always take us back to some non-explosive path. The path won't be steady, but our x_t won't blow up either. A parameter $|a| < 1$ dwarfs all forcing terms as long as $\sum_{s=-\infty}^{t} a^{t-s} b_s$ is finite. That is, as long as the b_t's don't grow at a rate faster than $1/a$.

Matters are totally different for $|a| \geq 1$. A homogeneous difference equation with solution $x_t^{compl.} = a^t x_0$ will explode. It will explode in a cycling pattern if a is negative, and it will explode in a straight pattern if a is positive. Similarly, any non-homogeneous difference equation will be explosive when $|a| \geq 1$. For any finite (or infinite) sum $\sum_{s=-\infty}^{t} a^{t-s} b_s$, $a^t x_0$ will make the x_t's blow up each period by an amount ax_0. So, there cannot be any form of stability. These concepts will soon carry over to higher-dimensional cases.

3 Linear first-order difference equations in n variables

3.1 The case of n variables

Economists often have to deal with difference equation systems that involve two or more variables. In principle, nothing changes. Just the names get more fancy, and the solution methods more tedious. For example, the higher order equivalents of the little coefficient a receive names such as 'eigenvalues' or 'characteristic roots' and are usually written as λ's (lambdas). In higher dimensions, these λ's can become complex numbers. Instead of asking whether they are bigger or smaller than one in absolute value (as is most appropriate for real numbers), people like to say these λ's lie inside or outside the unit circle— but they essentially mean the same. Pure mathematicians would probably prefer to say that the λ's are less or more than one in modu-

lus because the modulus for complex numbers is the equivalent to an absolute value for real numbers. Hence, higher-order systems will also be stable whenever these eigenvalues λ's are strictly less than one in modulus; the systems will be explosive as soon as one λ weakly exceeds one in modulus.

When solving first-order difference equations for n variables, the grand objective is to get back to the simplest possible case where $x_t = \lambda x_{t-1} + b_t$ as in (1) and the x's are scalars. Consider a first-order difference equation in n variables

$$\begin{pmatrix} x_t^1 \\ x_t^2 \\ \vdots \\ x_t^n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{n1} & \dots & & a_{nn} \end{pmatrix} \begin{pmatrix} x_{t-1}^1 \\ x_{t-1}^2 \\ \vdots \\ x_{t-1}^n \end{pmatrix} + \begin{pmatrix} b_t^1 \\ b_t^2 \\ \vdots \\ b_t^n \end{pmatrix}. \tag{7}$$

This is still a first-order system because only one lag of the vector $\mathbf{x}_t = (x_t^1, \dots, x_t^n)'$ appears. Suppose for a moment that we were in the most lovely of all cases: Suppose the big matrix $\mathbf{A}$ were a diagonal matrix with different (real or complex) entries on the diagonal but zeros everywhere else. Then we would be done. The big system would simply collapse to n equations of the form $x_t^i = a_{ii} x_{t-1}^i + b_t^i$. [Check.] And we know the solution to these: (2) or (3).

The second most lovely case would be one in which we come as close to the most lovely case as possible. Concretely, it is the case where $\mathbf{A}$ is as ugly as it may be, but we can find a closely related equation system

$$\hat{\mathbf{x}}_t = \Lambda \hat{\mathbf{x}}_{t-1} + \hat{\mathbf{b}}_t, \tag{8}$$

where Λ is a diagonal matrix and we can recover the original $\mathbf{x}_t = (x_t^1, \dots, x_t^n)'$ as $\mathbf{x}_t = \mathbf{P}\hat{\mathbf{x}}_t$, and the original $\mathbf{b}_t$ as $\mathbf{b}_t = \mathbf{P}\hat{\mathbf{b}}_t$ for some related matrix $\mathbf{P}$. This second most lovely case occurs almost always in economics (it requires the assumption that the eigenvalues are distinct). Clever mathematicians have worked it out. Some call the procedure 'diagonalizing' $\mathbf{A}$, others call it 'finding the eigenvalues', still others call it 'finding the characteristic roots' of $\mathbf{A}$. Before we consider the special case of two variables in a detailed example, let's get a first understanding of the general procedure.

The idea is the following. Under certain regularity conditions any $n \times n$ matrix $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}^{-1}$, where Λ is a diagonal $n \times n$ matrix and $\mathbf{P}$ has full rank. The entries on the diagonal of Λ are called the n

eigenvalues of $\mathbf{A}$, and the columns of $\mathbf{P}$ are the n according eigenvectors of $\mathbf{A}$. Knowing this, we can transform any first-order difference equation system in n variables, such as (7), in the following way.

First, find the eigenvalues of the matrix $\mathbf{A}$ and write them on the diagonal of another matrix, keeping all entries off the diagonal at zero. Call that matrix of eigenvalues Λ. Second, find the according eigenvectors and assemble them as the columns of the matrix $\mathbf{P}$. Third, pre-multiply $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{b}_t = \mathbf{P}\Lambda\mathbf{P}^{-1}\mathbf{x}_{t-1} + \mathbf{b}_t$ by $\mathbf{P}^{-1}$ to obtain $\hat{\mathbf{x}}_t = \Lambda\hat{\mathbf{x}}_{t-1} + \hat{\mathbf{b}}_t$, where $\hat{\mathbf{x}}_t \equiv \mathbf{P}^{-1}\mathbf{x}_t$, $\hat{\mathbf{x}}_{t-1} \equiv \mathbf{P}^{-1}\mathbf{x}_{t-1}$, and $\hat{\mathbf{b}}_t \equiv \mathbf{P}^{-1}\mathbf{b}_t$ for all t. Fourth, since Λ is a convenient diagonal matrix, the solutions to the system $\hat{\mathbf{x}}_t = \Lambda\hat{\mathbf{x}}_{t-1} + \hat{\mathbf{b}}_t$ are all well known from (2) and (3). Hence, for each entry $\hat{x}_t^i$ in the p-vector $\hat{\mathbf{x}}_t$, just use the general solution $\hat{x}_t^i = c_i \cdot (\lambda_i)^t + \sum_{s=-\infty}^{t}(\lambda_i)^{t-s}\hat{b}_s$ when $|a| < 1$ (and the according solution when $|a| > 1$). The λ_i is the i^{th} eigenvalue in Λ. Now you will need n boundary conditions to precisely pin down all c_i's. Finally, pre-multiply the solution $\hat{\mathbf{x}}_t$ by $\mathbf{P}$ again, in order to get the solutions to the original $\mathbf{x}_t$. You're done.

As will become clear in the following example with two variables soon, the stability of the system will depend on the eigenvalues. Whenever all eigenvalues are strictly less than one in modulus, the system will be stable. As soon as one eigenvalue weakly exceeds unity, the system becomes explosive.

3.2 The same but in 2 variables and for an autonomous system

Let's consider the case of $n = 2$ as an example. By way of doing so, we can also derive some general formulas for both the eigenvalues and the eigenvectors in the two-variable case. Our objective is to find a solution to the following autonomous first-order system of difference equations

$$\begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_{t-1}^1 \\ x_{t-1}^2 \end{pmatrix} + \begin{pmatrix} b^1 \\ b^2 \end{pmatrix}, \tag{9}$$

or $\mathbf{x}_t = \mathbf{A}_{2\times 2}\mathbf{x}_{t-1} + \mathbf{b}$. Again, if we were in the most lovely of all cases with $a_{21} = a_{12} = 0$, we could simply use the solutions from the one-variable case, (2) or (3). So, let's 'decouple' the system and get to that case, that is, let's find

$$\begin{pmatrix} \hat{x}_t^1 \\ \hat{x}_t^2 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \hat{x}_{t-1}^1 \\ \hat{x}_{t-1}^2 \end{pmatrix} + \begin{pmatrix} \hat{b}^1 \\ \hat{b}^2 \end{pmatrix}, \tag{10}$$

where the x_t^i and b^i can be recovered from the $\hat{x}_t^i$ and $\hat{b}^i$. We proceed in five steps and make use of the Superposition Principle. Over the course of the first four steps we will find the solution to the *homogeneous* difference equation $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1}$. Then, in a fifth and last step, we will derive a particular solution to the non-homogeneous but *autonomous* system.

First we go after the eigenvalues. By definition, the eigenvalues of an $n \times n$ matrix $\mathbf{A}$ are the n (possibly non-distinct) values of λ that satisfy the equation

$$\det[\mathbf{A} - \lambda \cdot \mathbf{I}_n] = 0, \tag{11}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\det[\cdot]$ denotes the determinant. In the 2×2 case, this means that

$$\det\left[\begin{pmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{pmatrix}\right] = 0.$$

Writing these conditions out, we find

$$(a_{11} - \lambda)(a_{22} - \lambda) - a_{21}a_{12} = \lambda^2 - \text{tr}(\mathbf{A})\lambda + \det(\mathbf{A}) = 0,$$

where $\text{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$. Hence, the two values of λ that satisfy (11) in the 2×2 case are the roots of a quadratic equation. In particular, they are

$$\boxed{\lambda_{1,2} = \frac{\text{tr}(\mathbf{A})}{2} \pm \frac{1}{2}\sqrt{[\text{tr}(\mathbf{A})]^2 - 4\det(\mathbf{A})}.} \tag{12}$$

So, in the 2×2 case we simply need to know the trace and the determinant of $\mathbf{A}$ and we are done.

Second, let's find the according eigenvectors. The i^{th} eigenvector of the matrix $\mathbf{A}$ is defined as the vector $\mathbf{e}_i$ that solves the equation

$$\mathbf{A}\mathbf{e}_i = \mathbf{e}_i\lambda_i. \tag{13}$$

Note that if we find one vector that satisfies (13), then we actually found a whole bundle of them. Since we can multiply both sides of (13) by any number and still satisfy it, we can multiply $\mathbf{e}_i$ by any number and still satisfy (13). Therefore, we can choose the eigenvector that we prefer. Many mathematicians like to find the special eigenvector that has length one in

Euclidean space. However, there is a much more convenient eigenvector to pick in the 2×2 case: The eigenvector that has a y-coordinate of one.[1] With this convenient standardization, we only need to find the x-coordinate of the eigenvector. That is, we only need to find e_i in

$$\mathbf{A} \cdot \begin{pmatrix} e_i \\ 1 \end{pmatrix} = \begin{pmatrix} e_i \\ 1 \end{pmatrix} \cdot \lambda_i.$$

This is a system in two equations, and can be written out as

$$\begin{aligned} a_{11}e_i + a_{12} &= \lambda_i e_i \\ a_{21}e_i + a_{22} &= \lambda_i. \end{aligned}$$

Two equations and only one unknown e_i? As it turns out, these two equations are always identical by virtue of λ_i being an eigenvalue. [Check.] Solving either one of the two equations, we find the eigenvector $\mathbf{e}_i$ that corresponds to the eigenvalue i:

$$\boxed{\mathbf{e}_i = \begin{pmatrix} \frac{\lambda_i - a_{22}}{a_{21}} \\ 1 \end{pmatrix}.} \tag{14}$$

Third, we can make use of these definitions for our (homogeneous) second-order difference equation $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1}$. For that, let's write the eigenvalues and the eigenvectors into two matrices, Λ and $\mathbf{P}$:

$$\Lambda \equiv \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \quad \text{and} \quad \mathbf{P} \equiv (\mathbf{e}_1, \mathbf{e}_2) = \begin{pmatrix} \frac{\lambda_1 - a_{22}}{a_{21}} & \frac{\lambda_2 - a_{22}}{a_{21}} \\ 1 & 1 \end{pmatrix}. \tag{15}$$

The order of the eigenvalues and eigenvectors does not matter, but it must be the same in both matrices Λ and $\mathbf{P}$. From the definition of the eigenvector (13) we know that we can write $\mathbf{AP} = \mathbf{P}\Lambda$. Now suppose that the two eigenvalues are distinct. Then we can invert $\mathbf{P}$ because it must have full rank for distinct eigenvalues [check], and post-multiply both sides of $\mathbf{AP} = \mathbf{P}\Lambda$ with $\mathbf{P}^{-1}$. Hence,

$$\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}^{-1}.$$

[1] It is permissible to rule out a y-coordinate of zero because we wouldn't have started the whole procedure if there had been a column vector with a zero y-coordinate in $\mathbf{A}$ in the first place.

This is perfect for our purposes. We can simply plug the result into our homogeneous difference equation system $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1}$ and get

$$\mathbf{x}_t = \mathbf{P}\Lambda\mathbf{P}^{-1}\mathbf{x}_{t-1}.$$

Pre-multiplying the entire thing by $\mathbf{P}^{-1}$ and defining $\hat{\mathbf{x}}_t \equiv \mathbf{P}^{-1}\mathbf{x}_t$ for all t, we have a transformed system that we can solve:

$$\tilde{\mathbf{x}}_t = \Lambda\hat{\mathbf{x}}_{t-1}.$$

We cannot only solve it, we even know the solutions from the scalar case, (2) or (3). Hence,

$$\hat{\mathbf{x}}_t = \begin{pmatrix} c_1(\lambda_1)^t \\ c_2(\lambda_2)^t \end{pmatrix}. \tag{16}$$

Fourth, we can recover our original variable $\mathbf{x}_t$ from $\hat{\mathbf{x}}_t$ as $\mathbf{x}_t = \mathbf{P}\hat{\mathbf{x}}_t$. So, finally, we have the desired general solution for our homogeneous first-order difference equation system in two variables, $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1}$:

$$\mathbf{x}_t = \begin{pmatrix} \frac{\lambda_1 - a_{22}}{a_{21}} & \frac{\lambda_2 - a_{22}}{a_{21}} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} c_1(\lambda_1)^t \\ c_2(\lambda_2)^t \end{pmatrix}. \tag{17}$$

The constants c_1 and c_2 need to be determined by two boundary conditions.

Fifth, if we finally find any particular solution to our original, autonomous difference equation $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{b}$, we are done. The Superposition Principle tells us that we just need to add such a particular solution and the solution to the homogeneous system, (17). Any autonomous linear difference equation system has a steady state (under some regularity condition). This steady state is the simplest particular solution that we can find. Try it. Suppose there is a steady state for which $\mathbf{x}_t = \mathbf{x}_{t-1} = \bar{\mathbf{x}}$. Then, the original difference equation becomes $\bar{\mathbf{x}} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{b}$. Hence, $(\mathbf{I}_n - \mathbf{A})\bar{\mathbf{x}} = \mathbf{b}$ and, in the 2×2 case,

$$\boxed{\bar{\mathbf{x}} = (\mathbf{I}_2 - \mathbf{A})^{-1}\mathbf{b} = \frac{1}{1 - \det(\mathbf{A}) - \mathrm{tr}(\mathbf{A})} \begin{pmatrix} 1 - a_{21} & a_{12} \\ a_{21} & 1 - a_{11} \end{pmatrix} \mathbf{b}.} \tag{18}$$

As long as $\det(\mathbf{A}) + \mathrm{tr}(\mathbf{A}) \neq 1$, the steady state exists. (So we have just proven to ourselves that any autonomous linear difference equation system in two variables has a steady state if $\det(\mathbf{A}) + \mathrm{tr}(\mathbf{A}) \neq 1$). Note that the term

$(\mathbf{I}_2 - \mathbf{A})^{-1}\,\mathbf{b}$ resembles the scalar case closely, where $\sum_{s=-\infty}^{t} a^{t-s}b = \frac{b}{1-a}$ (if $|a| < 1$).

We don't know whether the steady state is stable, that is, whether the economy will converge to it or not. This will depend on the eigenvalues. If at least one eigenvalue weakly exceeds unity in absolute value (or modulus), the steady state is unstable. That is, once we leave it, we can never return. If, on the other hand, all eigenvalues are strictly less than one in absolute value, the system is stable and we always converge back to the steady state no matter how far we were apart initially. In any case, an economy which starts out in the steady state will never leave it unless it is hit by an exogenous shock.

Using the Superposition Principle we can finally add the particular solution (18) (the steady state) to the complementary solution (17). This yields the general solution

$$\boxed{\mathbf{x}_t = \begin{pmatrix} \frac{\lambda_1 - a_{22}}{a_{21}} & \frac{\lambda_2 - a_{22}}{a_{21}} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} c_1(\lambda_1)^t \\ c_2(\lambda_2)^t \end{pmatrix} + \bar{\mathbf{x}},} \tag{19}$$

where the steady state $\bar{\mathbf{x}} = (\mathbf{I}_2 - \mathbf{A})^{-1}\mathbf{b}$ is given by (18) and the eigenvalues satisfy $\lambda_{1,2} = \frac{\mathrm{tr}(\mathbf{A})}{2} \pm \frac{1}{2}\sqrt{\mathrm{tr}(\mathbf{A})^2 - 4\det(\mathbf{A})}$ by (12). The coefficients c_1 and c_2 need to be determined by two boundary conditions.

With the solution (19), we can now make an argument about the stability of the system in more formal language: As can be seen from (19), the stability properties that we first encountered in section 2.2 directly carry over to the more general case of a system in two variables. Whenever at least one eigenvalue λ_i exceeds one in absolute value (or in modulus if λ_i is complex), our $\mathbf{x}_t$ will blow up. Consider the vector $(c_1(\lambda_1)^t, c_2(\lambda_2)^t)'$. Each period we add an increment $c_1\lambda_1$ to this vector. When $|\lambda_1| \geq 1$, this increment is non-decreasing and there is no way that we could find an $\mathbf{x}_t$ that converges to $\bar{\mathbf{x}}$. Right in the opposite, $\mathbf{x}_t$ will tend to infinity. When $|\lambda_1| < 1$ *and* $|\lambda_2| < 1$, however, the added increment is getting ever smaller over time. Thus, the vector $(c_1(\lambda_1)^t, c_2(\lambda_2)^t)'$ converges to a zero-vector and, as a result, the term $\mathbf{x}_t - \bar{\mathbf{x}}$ must go to zero. That is, $\mathbf{x}_t$ converges to the steady state level over time whenever *both* eigenvalues are less than one in modulus.

A non-autonomous system can be solved along similar lines, but the expressions for the forcing term will be more complicated, and there is no steady state in general.

4 Linear difference equations of order p

4.1 The general case

Following the same procedure, we will handle higher-order difference equations by conveniently reducing them to the cases before. As it will turn out, we have in fact already solved higher order difference equations. Consider any p^{th}-order difference equation in the scalar variable y_t

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + b_t. \tag{20}$$

Such an equation can always be rewritten as a *first*-order difference equation in p variables. Don't be surprised. This is almost trivial:

$$\begin{aligned}
\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t+1-p} \end{pmatrix} &= \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} b_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\
&= \mathbf{F}\cdot \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} b_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.
\end{aligned} \tag{21}$$

We simply write all the coefficients $\phi_1, \ldots, \phi_p$ into the first row of the matrix $\mathbf{F}$, plug ones into all entries below the diagonal and zeros elsewhere. The result is clearly a first-order system in the p-vector $\mathbf{y}_t \equiv (y_t, y_{t-1}, \ldots, y_{t+1-p})'$. By following precisely the procedure in section 3.1, we can find the eigenvalues and eigenvectors and solve (21) in just the same way as before.

This would even hold true for a p^{th}-order system in n variables. Suppose the n-vector $\mathbf{y}_t$ follows a p^{th}-order difference equation

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \ldots + \Phi_2 \mathbf{y}_{t-2} + \mathbf{b}_t,$$

where the Φ_i are $n \times n$ matrices, and $\mathbf{b}_t$ is an n-vector. Such an equation can always be rewritten in stacked form as

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t+1-p} \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_p \\ \mathbf{I}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \vdots & \ddots & & \vdots \\ \mathbf{0}_n & \cdots & \mathbf{I}_n & \mathbf{0}_n \end{pmatrix} \begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{pmatrix} + \begin{pmatrix} \mathbf{b}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix},$$

where $\mathbf{0}_n$ is an $n \times n$ matrix full of zeros, $\mathbf{0}$ is an n-vector of zeros, and $\mathbf{I}_n$ is the $n \times n$ identity matrix. Again, with the general method from section 3.1, we can solve this system.

4.2 The same but in 2 variables

Let's apply our findings to the case of a second order difference equation in a scalar variable y_t. As outlined above, the second-order difference equation

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + b_t \tag{22}$$

can be rewritten as a first-order difference equation system in two variables:

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} b_t \\ 0 \end{pmatrix}, \tag{23}$$

or $\mathbf{y}_t = \mathbf{F}\mathbf{y}_t + \mathbf{b}_t$. In section 3.2 we derived convenient formulas for a first-order difference equation in two variables. We first found the solution to the homogeneous difference equation. Using the Superposition Principle, we then added any particular solution of the non-homogeneous system and obtained the general solution. Let's follow the same procedure again. Formula (12) gives us the eigenvalues. Since $\text{tr}(\mathbf{F}) = \phi_1$ and $\det(\mathbf{F}) = -\phi_2$, we find

$$\boxed{\lambda_{1,2} = \tfrac{\phi_1}{2} \pm \tfrac{1}{2}\sqrt{\phi_1^2 + 4\phi_2}.} \tag{24}$$

You may have noted before that the eigenvalues of a 2×2 matrix are closely related to the roots of a quadratic equation. In the case of difference equations, the eigenvalues of the $p \times p$ matrix $\mathbf{F}$ are in fact equivalent to the p characteristic roots of the homogeneous difference equation. After all, a p^{th}-order homogeneous difference equation is nothing but a p^{th}-order polynomial. More on that in section 5.

The according eigenvectors are, by formula (14),

$$\boxed{\mathbf{e}_{1,2} = \begin{pmatrix} \lambda_{1,2} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{\phi_1}{2} \pm \frac{1}{2}\sqrt{\phi_1^2 + 4\phi_2} \\ 1 \end{pmatrix}.} \tag{25}$$

Finally, using result (17), we find the solution to the homogeneous second-order difference equation $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2}$

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} c_1(\lambda_1)^t \\ c_2(\lambda_2)^t \end{pmatrix},$$

or

$$\boxed{y_t = \hat{c}_1(\lambda_1)^t + \hat{c}_2(\lambda_2)^t.} \tag{26}$$

In (26), we conveniently defined $\hat{c}_i \equiv c_i\lambda_i$. We can always do that because the $\hat{c}_i$ will ultimately be determined correctly by the boundary conditions.

We can find the general solution to the non-homogeneous difference equation (22) by invoking the Superposition Principle once more. Provided that both eigenvalues are less than unity in modulus, the general solution is

$$\boxed{y_t = \hat{c}_1(\lambda_1)^t + \hat{c}_2(\lambda_2)^t + \frac{1}{(1-\lambda_1\mathbb{L})(1-\lambda_2\mathbb{L})}b_t,} \tag{27}$$

where $\frac{1}{(1-\lambda_1\mathbb{L})(1-\lambda_2\mathbb{L})}b_t$ is a particular solution to the second-order non-autonomous difference equation $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + b_t$. Why this is the case will become clear in the following section.

5 The equivalence of eigenvalues and characteristic roots in difference equations

This section is devoted to one sole goal: to show that all the fancy techniques that we applied to scalar difference equations in the preceding section 4 are no different from high school mathematics. We simply used more elaborate terms for it.

5.1 The general scalar case

Polynomials have characteristic roots. For example, quadratic equations such as $az^2 + bz + c = 0$ have two roots that solve them for z. Namely, $z_{1,2} = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$. Quadratic equations are second-order polynomials. Higher order polynomials have p such roots. There is an extremely powerful theorem in algebra, the Fundamental Theorem of Algebra. You may not have encountered this name for the theorem yet, but you most likely know it from high school, too. The Fundamental Theorem of Algebra tells us that any p^{th}-order polynomial can be factored into exactly p (possibly non-distinct) terms, and each of these terms involves one characteristic root. So, for any p^{th}-order polynomial in a real number z

$$z^p + d_1 z^{p-1} + d_2 z^{p-2} + \ldots + d_{p-1}z + d_p = 0$$

we can find p (possibly non-distinct) characteristic roots λ_1, λ_2, ... , λ_p. Using these characteristic roots we can rewrite the polynomial in equivalent terms as

$$\begin{aligned} & z^p + d_1 z^{p-1} + d_2 z^{p-2} + \ldots + d_{p-1} z + d_p = 0 \\ \Leftrightarrow \quad & (\lambda_1 - z)(\lambda_2 - z)(\lambda_3 - z) \cdots (\lambda_p - z) = 0. \end{aligned}$$

The p characteristic roots are functions of the parameters d_1, ... , d_p of the original polynomial.

This line of thought is often applied to difference equations, too. To see how, take our p^{th}-order scalar difference equation (20) in section 4.1,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + b_t,$$

retain the part of it that is a homogeneous difference equation, and rewrite that homogeneous difference equation as

$$(1 - \phi_1 \mathbb{L} - \phi_2 \mathbb{L}^2 - \ldots - \phi_p \mathbb{L}^p) \cdot y_t = 0 \tag{28}$$

in lag operators. Due to the properties of the lag operator, the term $(1 - \phi_1 \mathbb{L} - \phi_2 \mathbb{L}^2 - \ldots - \phi_p \mathbb{L}^p)$ in (28) can be considered equivalent to the following polynomial

$$(1 - \phi_1 z^{-1} - \phi_2 z^{-2} - \ldots - \phi_p z^{-p}) = 0. \tag{29}$$

It is equivalent in the sense that if we know how to factor this polynomial in z into p factors, we can also factor the term $(1 - \phi_1 \mathbb{L} - \phi_2 \mathbb{L}^2 - \ldots - \phi_p \mathbb{L}^p)$ into the same p factors, simply replacing the z's with $\mathbb{L}$'s. This may seem surprising at first. It really isn't. See the idea at work in the following paragraphs. [To convince yourself now, expand a second-order polynomial in z of the form $(1 - \lambda_1 z^{-1})(1 - \lambda_2 z^{-1})$ and a second-order polynomial in $\mathbb{L}$ of the form $(1 - \lambda_1 \mathbb{L})(1 - \lambda_2 \mathbb{L})$ to see that this equivalence is right. You can repeat that procedure for any higher-order polynomial.]

Unfortunately, we usually do not immediately know the characteristic roots of a polynomial of the form $(1 - \phi_1 z^{-1} - \phi_2 z^{-2} - \ldots - \phi_p z^{-p}) = 0$ where the z's are raised to negative powers. However, we do generally know the characteristic roots of a polynomial where the z's are raised to positive powers. So, let's simply multiply (29) through by z^p and obtain

$$z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \ldots - \phi_p = 0. \tag{30}$$

This expression is called the *associated polynomial* of the difference equation (20). By the Fundamental Theorem of Algebra, the *associated polynomial* can be factored into p terms of the form

$$\begin{aligned} & (\lambda_1 - z)(\lambda_2 - z)(\lambda_3 - z)\cdots(\lambda_p - z) \\ = \; & z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \ldots - \phi_{p-1} z - \phi_p \\ = \; & 0. \end{aligned} \tag{31}$$

The λ_i's are also called the characteristic roots of the difference equation.

Now let's take this finding back to our original p^{th}-order difference equation, written with lag operators in the form (28). We know that $(\lambda_1 - z) \cdot \cdots (\lambda_p - z) = 0$. Thus, in order to return to the first polynomial (29), we can multiply (31) by z^{-p} again and obtain

$$\begin{aligned} & 1 - \phi_1 z^{-1} - \phi_2 z^{-2} - \ldots - \phi_{p-1} z^{-p+1} - \phi_p z^{-p} \\ = \; & \left(\frac{\lambda_1}{z} - 1\right)\left(\frac{\lambda_2}{z} - 1\right)\left(\frac{\lambda_3}{z} - 1\right)\cdots\left(\frac{\lambda_p}{z} - 1\right) \\ = \; & 0. \end{aligned} \tag{32}$$

Simply divide each of the p factors by z. Now let's use the fact that the lag-operator $\mathbb{L}$ is behaving as if it were a real number z^{-1}. So, substitute $\mathbb{L}$ for λ in (32) and obtain

$$\begin{aligned} = \; & (\lambda_1 \mathbb{L} - 1)(\lambda_2 \mathbb{L} - 1)(\lambda_3 \mathbb{L} - 1)\cdots(\lambda_p \mathbb{L} - 1) \\ = \; & (1 - \lambda_1 \mathbb{L})(1 - \lambda_2 \mathbb{L})(1 - \lambda_3 \mathbb{L})\cdots(1 - \lambda_p \mathbb{L}) \\ = \; & 0. \end{aligned} \tag{33}$$

The equality holds because we can multiply each factor by -1 since the polynomial is equal to 0. And, after all, we have found an extremely handy expression for the original difference equation (20):

$$(1 - \lambda_1 \mathbb{L})(1 - \lambda_2 \mathbb{L})\cdots(1 - \lambda_p \mathbb{L}) \cdot y_t = b_t. \tag{34}$$

As mysterious as it may have looked initially, the lag operator in a polynomial behaves like the inverse of a real number. The last expression (34) may reconcile you. We know that we can re-write an expression $(1 - \lambda \mathbb{L}) \cdot y_t = b_t$ as $y_t = b_t / (1 - \lambda \mathbb{L}) = \sum_{s=-\infty}^{t} \lambda^{t-s} b_s$ if we have only one λ.[2] Expression (34)

[2] Strictly speaking, this infinite sum of past b_t only results if λ is strictly less than one in absolute value; an infinite sum of *future* b_t follows if λ is strictly bigger than one in absolute value.

re-iterates that insight for more than one lag of y_t. We can now write the expression $(1-\lambda_1\mathbb{L})\cdots(1-\lambda_p\mathbb{L})\,y_t = b_t$ as $y_t = b_t/\left[(1-\lambda_1\mathbb{L})\cdots(1-\lambda_p\mathbb{L})\right]$. So, y_t is nothing but a complicated summation of the past b_t.

We have called the p characteristic roots of the p^{th}-order difference equation lambdas as well. For a good reason. As it turns out, these $\lambda_1, \ldots, \lambda_p$ are exactly equal to the p eigenvalues of the difference equation (20), and its vector version (21), in section 4.1. In general:

> Factoring the p^{th}-order polynomial $(1-\phi_1\mathbb{L}-\phi_2\mathbb{L}^2-\ldots-\phi_p\mathbb{L}^p)\cdot y_t$ as
>
> $$(1-\lambda_1\mathbb{L})\,(1-\lambda_2\mathbb{L})\cdots(1-\lambda_p\mathbb{L})\cdot y_t$$
>
> is equivalent to finding the p eigenvalues $\lambda_1, \ldots, \lambda_p$ of the matrix $\mathbf{F}$ in the vector version of the difference equation,

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t+1-p} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} b_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \mathbf{F}\cdot \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} b_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Rather than proving the above statement in general terms, let's turn to the more intuitive case of a second-order difference equation in the next section. The above result is somewhat ironic. It tells us, in fact, that all the complicated sounding concepts such as eigenvalues and characteristic roots are nothing but what we used to derive in high-school exercises. We are still not doing anything different but factoring polynomials. We have merely found more fancy applications.

5.2 The same but for a second-order difference equation

Take a second-order difference equation

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + b_t$$

as in (22). This difference equation can be rewritten with lag operators as

$$\left(1 - \phi_1 \mathbb{L} - \phi_2 \mathbb{L}^2\right) y_t = b_t. \tag{35}$$

We want to find the characteristic roots of this process in order to see whether it is stable or not. (It is stable iff both characteristic roots are strictly less than one in modulus.) For this purpose, we want to factor the expression $(1 - \phi_1 \mathbb{L} - \phi_2 \mathbb{L}^2)$ into two terms so that

$$\left(1 - \lambda_1 \mathbb{L}\right)\left(1 - \lambda_2 \mathbb{L}\right) y_t = b_t. \tag{36}$$

We know from section 5.1 above that the expression $(1 - \phi_1 \mathbb{L} - \phi_2 \mathbb{L}^2) = 0$ is behaved like a polynomial of the form $(1 - \phi_1 z^{-1} - \phi_2 z^{-2}) = 0$. This looks a little ugly, but we need not worry because we know the solutions to the *associated polynomial*

$$z^2 - \phi_1 z - \phi_2 = 0. \tag{37}$$

From our quadratic formula we remember that the two roots solving (37) are

$$z_{1,2} = \lambda_{1,2} = \frac{\phi_1 \pm \sqrt{{\phi_1}^2 + 4\phi_2}}{2}. \tag{38}$$

From the general derivations in section 5.1 (see (32) and 33) we know that these characteristic roots $\lambda_{1,2}$ are the characteristic roots in the expression $\left(1 - \lambda_1 \mathbb{L}\right)\left(1 - \lambda_2 \mathbb{L}\right) y_t$. And we were looking for these roots. Hence, the second-order difference equation (22) can be factored as

$$\begin{aligned} &\left(1 - \frac{1}{2}\left(\phi_1 - \sqrt{{\phi_1}^2 + 4\phi_2}\right)\mathbb{L}\right) \times \\ &\quad \left(1 - \frac{1}{2}\left(\phi_1 + \sqrt{{\phi_1}^2 + 4\phi_2}\right)\mathbb{L}\right) y_t \;\; = \;\; b_t. \end{aligned} \tag{39}$$

We know even more. The characteristic roots of this process will either both be real (whenever $\phi_2 \geq -\phi_1^2/4$) or both complex (whenever $\phi_2 < -\phi_1^2/4$). Suppose they are real. When will the eigenvalues be less than unity in absolute value? We can derive some conditions. One is: If ϕ_2 is positive, then at least one eigenvalue exceeds unity iff $|\phi_1| > 1$. To see this, consider the case where ϕ_2 is zero. Then, the roots in (38) simplify to ϕ_1 and zero. Now consider ϕ_2 bigger than zero. Then one root will exceed $|\phi_1|$ and the other

root will fall below zero if ϕ_1 is positive (and above zero if ϕ_1 is negative). Hence, whenever ϕ_2 is positive, at least one eigenvalue exceeds unity if and only if ϕ_1 exceeds unity in absolute value. [You can prove some further relationships along these lines.]

Finally, let's think about the statement that the characteristic roots of the associated polynomial $\lambda_{1,2}$ are equivalent to the eigenvalues of the matrix $\mathbf{F}$ in the system

$$\begin{aligned}\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} &= \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} b_t \\ 0 \end{pmatrix} \\ &= \mathbf{F} \cdot \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} b_t \\ 0 \end{pmatrix}.\end{aligned}$$

We have derived the eigenvalues of $\mathbf{F}$ in section 4.2. Browsing back to formula (24), we find these eigenvalues to be

$$\lambda_{1,2} = \frac{\phi_1 \pm \sqrt{{\phi_1}^2 + 4\phi_2}}{2}.$$

No surprise, they are exactly the same as the characteristic roots in (38) above. So, again, we didn't do anything more fancy in section 4.2 than solve the simple quadratic equation

$$\lambda^2 - \phi_1 \lambda - \phi_2 = 0$$

for λ.

6 Autoregressive Processes

6.1 The general scalar case

An autoregressive process of order p is defined as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + c + \tilde{\epsilon}_t. \tag{40}$$

It is called autoregressive because y_t depends on its own lagged values (*auto* means 'self' in Latin) in a linear manner (regressively). This expression looks very familiar, doesn't it? It is in fact nothing but an old known, a linear difference equation of order p as we saw it before in (20). There is only

one difference. The new 'forcing term' $c + \tilde{\epsilon_t}$ consists of two components. The coefficient c is just a constant real number. The other component is the crucial one: $\tilde{\epsilon_t}$ is taken as a random variable now.

In all autoregressive processes the error terms, or disturbances, $\tilde{\epsilon_t}$ are assumed to obey certain requirements on their distribution. There are three layers of these requirements. The first layer is the least restrictive and is *always* assumed to be satisfied. We want the errors $\tilde{\epsilon_t}$ to have a constant mean of zero, a constant and finite variance, and a constant autocovariance of zero. In statistical notation, we want

$$\mathbb{E}\left[\tilde{\epsilon_t}\right] = 0 \quad \forall t, \tag{41}$$

$$\mathbb{V}\mathrm{ar}(\tilde{\epsilon_t}) = \mathbb{E}\left[\tilde{\epsilon_t}^2\right] = \sigma_\epsilon^2 \quad \forall t, \tag{42}$$

and

$$\mathbb{C}\mathrm{ov}(\tilde{\epsilon_t}, \tilde{\epsilon}_{t-s}) = \mathbb{E}\left[\tilde{\epsilon_t}\tilde{\epsilon}_{t-s}\right] = 0 \quad \forall t, s \neq t \tag{43}$$

for all our $\tilde{\epsilon_t}'$s.

Sometimes we might prefer to have the error terms not only uncorrelated, but even independent across time. That is, we want to strengthen (43) and make the $\tilde{\epsilon_t}'$s independent and identically distributed error terms

$$\tilde{\epsilon_t} \sim i.i.d., \tag{44}$$

where (41) and (42) are still assumed to hold. This kind of distribution of the error terms is called *white noise*, and is the second layer of requirements. We finally reach a third layer of requirements when we assume that the $\tilde{\epsilon_t}'$s are not only independent and identically distributed, but that they are normal random variables. That is

$$\tilde{\epsilon_t} \overset{iid}{\sim} N(0, \sigma_\epsilon^2) \quad \forall t. \tag{45}$$

This kind of distribution is often called *Gaussian white noise* after Gauss, the 'inventor' of the normal distribution.

As it turns out, none of these layers of distributions of the disturbances $\tilde{\epsilon_t}$ matters for the main properties of an autoregressive process such as (40)

above. That's why it's so nice to have these layers satisfied.[3] The key property of the autoregressive process is its so-called *stationarity*. In order to work with autoregressive processes in a sensible manner, we need stationarity to be satisfied. As it will turn out soon, stationarity is intimately tied to the stability of the according homogeneous difference equation. So we are back to our standard analysis of eigenvalues.

There are various forms of stationarity. One is called covariance (or weak) stationarity. Another one is called strong stationarity. Before considering the concepts more formally in section 6.2, let's move a little ahead and simply state a fact about weak stationarity in terms that are known to us. The statement is the following:

> An autoregressive process of order p,
>
> $$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + c + \tilde{\epsilon}_t,$$
>
> is *covariance stationary* if the related homogeneous difference equation of order p,
>
> $$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p},$$
>
> is stable. That is, whenever *all* p eigenvalues (or characteristic roots) of the homogeneous difference equation are less than one in modulus.

So, what we want to do is have a look at the eigenvalues (or characteristic roots) of the homogeneous difference equation

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p}.$$

In section 4.1 we have seen how we can find the eigenvalues of any such homogeneous difference equation. We have also seen that the system is explosive as soon as one eigenvalue weakly exceeds one in modulus. On the other

[3] It can be shown that the following statements hold for more complicated processes of the errors, too. In particular, any invertible so-called ARMA process can be transformed into a pure autoregressive process of infinite order (AR(∞)). One extremely convenient property about polynomials is that infinite-order polynomials are not essentially different from finite polynomials. Hence, very similar concepts as the ones discussed here apply in general. The AR(∞) representation of any invertible ARMA process is stationary if all of the infinitely many characteristic roots are less than one in modulus.

hand, the system is stable as long as all eigenvalues are strictly less than unity. So, whenever all characteristic roots of the according homogeneous difference equation are strictly less than one in modulus, the autoregressive process (40) is *covariance stationary*.

6.2 Covariance stationarity

So far we have tied together quite substantial things. We have seen that eigenvalues of difference equations are nothing but the characteristic roots of certain polynomials as we know them from high-school. We have also seen that we can factor any difference equation using lag operators and these characteristics roots. Let's finally embed covariance stationarity in this framework.

We still need a good definition of stationarity. For the purpose of this section, covariance stationarity proves to be particularly well suited. It is also the concept of stationarity that most time-series econometricians mean when they talk about stationarity. Covariance stationarity, also called weak stationarity, is less restrictive than strong stationarity. It is therefore also the more natural concept to impose on stochastic process such as our p^{th}-order autoregressive process

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + c + \tilde{\epsilon}_t$$

in (40).

A serially correlated random variable y_t is a random variable that is correlated with its own past realizations. Our autoregressive y_t in (40) is an example of a serially correlated random variable. The mean and the variance of a serially correlated random variable are nothing special, simply $\mathbb{E}[y_t]$ and $\mathbb{V}\text{ar}(y_t)$. The so-called autocovariance is the covariance of y_t with its own lagged realizations. That is, the autocovariance of y_t is $\mathbb{C}\text{ov}(y_t, y_{t-s}) = \mathbb{E}[(y_t - \mathbb{E}[y_t]) \cdot (y_{t-s} - \mathbb{E}[y_{t-s}])]$. Thus, $\mathbb{V}\text{ar}(y_t) = \mathbb{C}\text{ov}(y_t, y_{t-s})$ iff $s = 0$. With this at hand, we can state the definition of covariance stationarity for a stochastic process y_t.

> A serially correlated random variable y_t is said to be covariance stationary if, at all times t,
>
> $$\mathbb{E}[y_t] = \mu$$

and

$$\mathbb{E}\left[\left(y_t-\mu\right)\left(y_{t-s}-\mu\right)\right]=\begin{cases}\gamma(s) & \text{for } s\neq 0\\ \sigma_y^2 & \text{for } s=0\end{cases}$$

are satisfied.

That is, both the mean and the variance of y_t must be time-independent and finite. In addition, the autocovariance must be finite and only a function of the time-lag s. The autocovariance must not depend on the time t at which we look at it. Note that, by the properties of the expectations operator $\mathbb{E}[\cdot]$, $\gamma(s)=\gamma(-s)$. [Check.]

Let's see how our p^{th}-order autoregressive process (40) fits into the picture. For now, let's assume that it satisfies one of the three requirement of stationarity and see how far we get. Let's pretend that the mean of y_t is finite and time-independent: $\mathbb{E}[y_t]=\mu$. Then, taking expectations of both sides of (40) yields

$$\mathbb{E}[y_t]=\phi_1\mathbb{E}[y_{t-1}]+\phi_2\mathbb{E}[y_{t-2}]+\ldots+\phi_p\mathbb{E}[y_{t-p}]+c.$$

Under the assumption that y_t is stationary, this expression simplifies to

$$\mu=\phi_1\mu+\phi_2\mu+\ldots+\phi_p\mu+c,$$

or

$$\mathbb{E}[y_t]=\mu=\frac{c}{1-\phi_1-\phi_2-\ldots-\phi_p}. \tag{46}$$

So, if it is stationary, we can conveniently rewrite the autoregressive process (40) as

$$y_t-\mu=\phi_1\left(y_{t-1}-\mu\right)+\phi_2\left(y_{t-2}-\mu\right)+\ldots+\phi_p\left(y_{t-p}-\mu\right)+\tilde{\epsilon}_t. \tag{47}$$

With this expression, it is particularly simple to derive the autocovariance of y_t given that it is stationary. We simply multiply (47) by $y_{t-s}-\mu$ and take expectations:

$$\begin{aligned}\gamma(s) &= \mathbb{C}ov\left(y_t,y_{t-s}\right)=\mathbb{E}\left[\left(y_t-\mu\right)\cdot\left(y_{t-s}-\mu\right)\right]\\ &= \mathbb{E}[\left(\phi_1\left(y_{t-1}-\mu\right)+\ldots+\phi_p\left(y_{t-p}-\mu\right)+\tilde{\epsilon}_t\right)\cdot\left(y_{t-s}-\mu\right)]\\ &= \phi_1\mathbb{E}\left[\left(y_{t-1}-\mu\right)\left(y_{t-s}-\mu\right)\right]+\ldots+\phi_p\mathbb{E}\left[\left(y_{t-p}-\mu\right)\left(y_{t-s}-\mu\right)\right]\\ &= \phi_1\gamma(s-1)+\ldots+\phi_p\gamma(s-p) \qquad (48)\end{aligned}$$

for $s \neq 0$. In the step from the second to the third line we made use of the fact that $\mathbb{E}\left[\tilde{\epsilon}_t \left(y_{t-s} - \mu\right)\right] = 0$ since y_{t-s} can only be a function of all the shocks that have occured up to period $t - s$ and, by assumption (43), the disturbances are not correlated over time.

The fact that (48) holds for any autoregressive process, so that $\gamma_s = \phi_1 \gamma_{s-1} + \ldots + \phi_p \gamma_{s-p}$, is an extremely interesting result. It says that the autocovariance of y_t, $\gamma(s)$, follows a homogeneous difference equation of order p. But not only that. The autocovariance $\gamma(s)$ follows exactly the same homogeneous difference equation as y_t itself! The result is not only interesting. It is highly convenient, too. We know under what conditions a p^{th}-order homogeneous difference equation is stable. It is stable if and only if all the characteristic roots of the associated polynomial are less than one in modulus. So, first of all, $\gamma(s)$ is indeed only a function of the lag s and not of time t. Second, it is finite if and only if the p characteristic roots of the homogeneous difference equation $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p}$ are all less than one in modulus. Hence, given our initial assumption that $\mathbb{E}\left[y_t\right] = \mu$ is not time-dependent, we can infer that the second requirement on stationarity is satisfied.

So, let's consider the third requirement for stationarity, that the variance of y_t be time-independent and finite. Multiplying (47) by $y_t - \mu$, taking expectations, and making use of the fact that $\mathbb{E}\left[\tilde{\epsilon}_t \left(y_{t-1} - \mu\right)\right] = 0$, yields

$$\begin{aligned} \mathbb{V}ar\left(y_t\right) &= \mathbb{E}\left[\left(\phi_1 \left(y_{t-1} - \mu\right) + \ldots + \phi_p \left(y_{t-p} - \mu\right) + \tilde{\epsilon}_t\right)\left(y_t - \mu\right)\right] \\ &= \phi_1 \gamma(1) + \ldots + \phi_p \gamma(p) + \sigma_\epsilon^2, \end{aligned} \tag{49}$$

[Check.] Note that the variance is certainly time-independent if the autocovariances $\gamma(1)$, $\ldots$, $\gamma(p)$ are. Since σ_ϵ^2 is finite by (42), and the coefficients ϕ_1, $\ldots$, ϕ_p are finite, all depends on these autocovariances. But we just saw in the paragraph before that these autocovariances are finite whenever $\mathbb{E}\left[y_t\right] = \mu$ and the eigenvalues of the homogeneous difference equation $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p}$ are less than one in modulus. So, the last thing we need to show is that $\mathbb{E}\left[y_t\right] = \mu$ is finite whenever the eigenvalues are less than one in modulus. Let's do that in the concluding section. Before we finally close this all, let's get some more intuition and consider a particular example where covariance stationarity fails.

6.3 A simple non-stationary process

A particularly neat example of failing covariance stationarity is the *random walk*. A random-walk is a first-order difference equation with a unit root:

$$y_t = y_{t-1} + \tilde{\epsilon}_t. \tag{50}$$

This is clearly a first-order difference equation. In fact, it is nothing but our very first difference equation (1) in a specific new form. We have just replaced the x with a variable y, have assumed that $a = 1$, and have made the forcing term stochastic. Remember that the little coefficient a is exactly what we later preferred to call the eigenvalue of a difference equation. When we rewrite the difference equation (50) in lag operators as

$$(1 - a\mathbb{L})y_t = \tilde{\epsilon}_t \tag{51}$$

this becomes even more clear. The little a is playing the role of the eigenvalue. As soon as the little a takes the unpleasant value of one, things don't work any more. First of all, we cannot divide (51) through by $(1 - \mathbb{L})$ any longer because, by the rules of the lag operator, we are only allowed to do so if $|a| < 1$. (Using lead operators won't help either since $\frac{1}{a} = 1$, too.) Second, y_t will not be stationary any more. To see this, let's first go back to our most primitive tool: recursion. We only applied it once in all these five sections, that was in (5) in section 2. This doesn't say that recursion has not been of much use. Right in the opposite, all our subsequent derivations depended on this early recursion argument. Let's use recursion once more:

$$y_t = y_{t-1} + \tilde{\epsilon}_t = y_{t-2} + \tilde{\epsilon}_{t-1} + \tilde{\epsilon}_t = \ldots = \sum_{s=-\infty}^{t} \tilde{\epsilon}_s. \tag{52}$$

So, if y_t follows a random walk, none of the past shock loses importance over time. In fact, y_t is a simple, unweighted sum of all past shocks, and an $\tilde{\epsilon}_t$ realization at a time close to the infinite past matters just as much for y_t today as did the shock yesterday. This is very different to the case of a first-order autoregressive process where the a is (just a tiny little bit) less than one. For $y_t = .99999 \cdot y_{t-1} + \tilde{\epsilon}_t$, yesterday's shock would matter a tiny little bit more than the shock two periods ago, but a whole lot more than the shock 100.000 periods ago which only receives a weight of $(.99999)^{100.000} = .36788$ for y_t 100.000 periods later. This shows that there is an essential difference

between $a = 1$ and $a = .99999$ even though they may be indistinguishable in any real data.

To make covariance stationarity fail, it is enough that one of the three requirements in the definition (p. 22) is violated. The easiest requirement to check is the variance of y_t. Without loss of generality suppose that $y_0 = 0$. That is, suppose there was a period in this infinitely long lasting stochastic process, call it $t = 0$, in which all past observations of $\tilde{\epsilon}_t$, namely $\tilde{\epsilon}_{-1}$, $\tilde{\epsilon}_{-2}$, ... happened to cancel each other. Now, to show that a random walk is not covariance stationary, let's go for a proof by contradiction. Suppose a random walk were covariance stationary. Then $\mathbb{V}ar(y_0)$ must be finite. With this, we can consider the variance of y_1 one period after the period in which all past $\tilde{\epsilon}_t$ miraculously canceled without loss of generality.

$$\mathbb{V}ar(y_1) = \mathbb{V}ar(y_0 + \tilde{\epsilon}_1) = \mathbb{V}ar(y_0) + \sigma_\epsilon^2,$$

where the last step follows from the fact that the disturbances $\tilde{\epsilon}_t$ are not correlated over time (recall requirement (43)). By the same token, we can also take the variance of y_T at some future time $T > 1$. By recursion, we soon find

$$\mathbb{V}ar(y_T) = \mathbb{V}ar\left(y_0 + \sum_{s=1}^{T} \tilde{\epsilon}_s\right) = \mathbb{V}ar(y_0) + T \cdot \sigma_\epsilon^2,$$

where the last step again follows from the fact that the disturbances $\tilde{\epsilon}_t$ are not correlated over time. [Check.] Clearly, the variance is time-dependent now. Not only that, it is even tending to infinity as T goes to infinity. We have successfully contradicted our initial assertion that y_t is stationary. We cannot even start under the (wrong) assumption that $\mathbb{V}ar(y_0)$ is finite. $\mathbb{V}ar(y_0)$ is as infinite at $t = 0$ as it is in any other period.

What can we do with such a nasty process to make it well behaved? We can domesticate it. Time-series econometricians call this domestication procedure 'differencing'. By 'differencing' they mean that you should pre-multiply your original (and nasty) process by a difference term $(1-\mathbb{L})$. Doing that in the case of our random walk (50), we find

$$(1 - \mathbb{L})y_t = y_t - y_{t-1} = \tilde{\epsilon}_t.$$

Thus, the first difference of a random walk y_t follows the same process as $\tilde{\epsilon}_t$. However, $\tilde{\epsilon}_t$ is stationary by assumptions (41), (42), and (43). [Check.] So,

differencing makes nasty processes stationary again. In the case of a random walk, differencing once resolves the problem. In even nastier cases where processes have more than one unit root, repeated differencing will take care of the problem.

We have seen that unit roots make stochastic processes explosive in the sense that the variance of the serially correlated random variable becomes infinite. Similarly, roots that *exceed* one in absolute value will cause non-stationarity. By another nice property of polynomials, repeated differencing will even make processes with roots that exceed one in modulus stationary again.

6.4 Stationarity of a finite autoregressive process

One thing remains to show from section 6.2: that the mean of a p^{th}-order stochastic process is constant and finite if all characteristic roots of the associated polynomial are less than one in modulus. All arguments in section 6.2 were based on this assumption, $\mathbb{E}[y_t] = \mu$ for all t. If we can prove this, we have finally shown that characteristic roots below one in modulus guarantee covariance stationarity. We will have proven beyond reasonable doubt that

> An autoregressive process of order p,
>
> $$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + c + \tilde{\epsilon}_t,$$
>
> is *covariance stationary* if the related homogeneous difference equation of order p,
>
> $$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p},$$
>
> is stable. That is, whenever *all* p eigenvalues (or characteristic roots) of the homogeneous difference equation are less than one in modulus.

Let's go. Take expectations of both sides of (40) to find

$$\mathbb{E}[y_t] = \phi_1 \mathbb{E}[y_{t-1}] + \phi_2 \mathbb{E}[y_{t-2}] + \ldots + \phi_p \mathbb{E}[y_{t-p}] + c.$$

Suppose the expectations were time-dependent indeed. Then we can write this difference equation in lag operators as

$$\begin{aligned} & (1 - \phi_1 \mathbb{L} - \phi_2 \mathbb{L}^2 - \ldots - \phi_p \mathbb{L}^p) \mathbb{E}[y_t] = c \\ \iff \; & (1 - \lambda_1 \mathbb{L})(1 - \lambda_2 \mathbb{L}) \cdots (1 - \lambda_p \mathbb{L}) \mathbb{E}[y_t] = c, \end{aligned} \tag{53}$$

where we made use of the derivations in section 2.2, of (32) and (33) in particular. So, if all eigenvalues $\lambda_1, \ldots, \lambda_p$ are less than one in modulus, we can bring the factors $(1 - \lambda_1\mathbb{L}), \ldots, (1 - \lambda_p\mathbb{L})$ in (53) over to the other side one by one. Consider the first one. Divide both sides of (53) by $(1 - \lambda_1\mathbb{L})$. This yields $\frac{c}{1-\lambda_1\mathbb{L}} = \frac{c}{1-\lambda_1}$ since c is a finite constant. Repeat this with the next one to find $\frac{c}{(1-\lambda_2\mathbb{L})(1-\lambda_1)} = \frac{c}{(1-\lambda_2)(1-\lambda_1)}$, and so forth. Finally, we obtain

$$\mathbb{E}[y_t] = \frac{c}{(1-\lambda_1)(1-\lambda_2)\cdots(1-\lambda_p)} = \frac{c}{1-\phi_1-\phi_2-\ldots-\phi_p}.$$

Hence, $\mathbb{E}[y_t]$ is both finite and constant whenever c is constant and all eigenvalues $\lambda_1, \ldots, \lambda_p$ are less than one in modulus.

That's it.

We have shown to ourselves: The eigenvalues of the matrix in a system of linear difference equations determine the stability of this system. The system is explosive when at least one eigenvalue exceeds unity. In the case of scalar difference equations, these very same eigenvalues can also be interpreted as the characteristic roots of an associated polynomial. So, in fact, we don't do anything more sophisticated but factor polynomials when we solve difference equations. Finally, when time-series econometricians talk about unit roots or roots outside the unit circle, they don't say anything new either. They just solve a polynomial and tell us that a stochastic process is explosive when these roots are bigger than unity.