Mechanics of Forming and Estimating Dynamic Linear Economies

Lars Peter Hansen, Ellen R. McGrattan,
and Thomas J. Sargent

Introduction

This paper describes how recursive linear control and estimation theory can be applied to estimate dynamic equilibrium models. Recursive linear control theory can be used to compute equilibria of linear-quadratic economies and linearly to approximate solutions of nonlinear economies. Equilibrium conditions define a mapping from a model's free parameters, describing preferences, technologies, endowments, information, and government policies, to equilibrium stochastic processes of observable variables. The estimation problem is roughly speaking to invert that mapping and to use time series of observations on some of the variables in the model to make inferences about the model's free parameters in light of the mapping defining the equilibrium stochastic process. Maximum likelihood and the method of moments are used to extract parameter estimates from time series data. Recursive linear estimation theory can be used to compute a Gaussian likelihood function.[1] This paper describes a collection of procedures for speedily calculating equilibria, for computing an approximate likelihood function, and for maximizing that likelihood function. The duality of linear control and filtering theory imparts a unity to these procedures.[2]

Among the conveniences afforded by this framework is the ability analytically to differentiate the likelihood function with respect to the free parameters of the economic model. Obtaining these derivatives involves, via a chain rule, two differentiations of solutions of some Riccati equations with respect to the parameters in their return (or covariance) and transition matrices. First, we must differentiate the equilibrium with respect to its free parameters; and second, we must differentiate the parameters of the "innovations representation" or "vector autoregression" with respect to parameters of measurement error processes and the equilibrium stochastic process for the economic model. It is the relative ease of accomplishing the second piece of the job that makes linear-quadratic models especially convenient. We describe the nuts and bolts of these calculations.

This paper is organized as follows. We display two types of economies and how they are associated with social planning problems that can be formu-

[1] Parts of this paper rely heavily on Anderson and Moore (1970, pp. 158–161). For general background, see Kwakernaak and Sivan (1972) or Sargent (1980). The former mostly treats continuous time systems, while the latter focuses on discrete time systems.

[2] Duality refers to the applicability of identical mathematics to solve the classical control and filtering problems.

lated as optimal linear regulator problems. We describe the optimal linear regulator, then display two tricks of the trade, namely, a pair of transformations that remove both discounting and cross-products between states and controls. Next we describe Vaughan's (1970) eigenvector method for solving an optimal linear regulator problem without iterating on Bellman's equation. Vaughan's method is typically much faster than Bellman's. We describe how Vaughan's algorithm can be used to compute an equilibrium for a distorted economy. As an alternative to Vaughan's method, we can use a closely related method called the *doubling algorithm*, which we explain next. We then show how the calculations can be further accelerated by partitioning the state vector to achieve a "controllability canonical form." We describe how to use the Kalman filter to obtain an innovations representation and how to use it to compute a Gaussian likelihood function. Finally, we display formulas for the gradient of the log of Gaussian likelihood function with respect to free parameters of an economic model. These formulas are homely, but easy to program and useful for accelerating the process of maximizing the likelihood function.

Two Economies

General strategies

A class of asset pricing and real business cycle models uses the *optimal linear regulator* problem as the workhorse for computing equilibria. After an equilibrium has been computed, the Kalman filter can be used to deduce the vector autoregressive representation for variables that are linear functions of the state. The autoregressive representation is used to interpret the data, either informally or to form the Gaussian likelihood function recursively.

Two general types of models are used, which differ with respect to the point in the analysis at which linear-quadratic approximations are imposed or how they are interpreted. In the first type of model, preferences are specified to be quadratic functions and transition laws are linear ones. The second type of model uses a linear regulator problem to approximate a dynamic programming problem that is not itself linear-quadratic.

Linear-quadratic economy

There is an exogenous information vector z_t governed by

$$z_{t+1} = A_{22} z_t + C_2 w_{t+1}, \tag{1}$$

where w_{t+1} is a martingale difference sequence with $E w_t w_t' = I$, and the eigenvalues of A_{22} are bounded in modulus by $1/\sqrt{\beta}$. The vector z_t determines a preference shock process b_t and an endowment shock process d_t via

$$\begin{aligned} d_t &= U_d \ z_t \\ b_t &= U_b \ z_t . \end{aligned} \tag{2}$$

A representative household has preferences ordered by

$$E_0 \sum_{t=0}^{\infty} \beta^t (s_t - b_t) \cdot (s_t - b_t), \quad 0 < \beta < 1, \tag{3}$$

where s_t is a vector of household services produced at time t via the household technology

$$\begin{aligned} s_t &= \Lambda h_{t-1} + \Pi c_t \\ h_t &= \Delta_h h_{t-1} + \Theta_h c_t, \end{aligned} \tag{4}$$

where h_t is a vector of household durable goods at t, c_t is a vector of rates of consumption, and Λ, Π, Δ_h, Θ_h are matrices with the eigenvalues of Δ_h bounded in modulus by $1/\sqrt{\beta}$.

There is a constant returns to scale production technology

$$\begin{aligned} \Phi_c c_t + \Phi_i i_t &= \Gamma k_{t-1} + d_t \\ k_t &= \Delta_k k_{t-1} + \Theta_k i_t, \end{aligned} \tag{5}$$

where k_t is a vector of capital goods used in production, i_t is a vector of investment goods, and Δ_k is a matrix whose eigenvalues are bounded in modulus by $1/\sqrt{\beta}$.

The social planning problem in this economy is to maximize (3) over choices of contingency plans for $\{c_t, i_t, k_t, h_t\}_{t=0}^{\infty}$ subject to (1), (2), (4), and (5) and subject to given initial conditions for (z_0, h_{-1}, k_{-1}). The social planning problem fits within the optimal linear regulator framework and leads to a quadratic optimal value function $V(x_0) = x_0' P x_0 + \rho$ where $x_t' = [h_{t-1}, k_{t-1}, z_t]$. The law of motion for the economy is of the form

$$x_{t+1} = A_o x_t + C w_{t+1}.$$

Hansen and Sargent (1994) describe a competitive equilibrium for this economy. Scaled time 0 Arrow-Debreu prices of the consumption vector denoted p_t^0 can be computed from the information in (P, A_o) and the household technology parameters and turn out to be a linear function of the state:

$$p_t^0 = M_c x_t / \mu_0^w,$$

where M_c is a matrix and μ_0 is a positive scalar giving the numeraire or marginal utility of wealth.

The price of a claim to a stream of consumption vectors $\xi_t = S_\xi x_t$ is given by

$$a_0 = E_0 \sum_{t=0}^{\infty} \beta^t p_t^0 \cdot \xi_t$$

or

$$a_0 = E \sum_{t=0}^{\infty} \beta^t \, x_t' \, Z_a x_t \mid I_0, \tag{6}$$

where

$$Z_a = S'_\xi M_c / \mu_0^w . \tag{7}$$

Hansen and Sargent show that a_0 can be represented as

$$a_0 = x'_0 \, \mu_a \, x_0 + \sigma_a , \tag{8}$$

where

$$\mu_a = \sum_{\tau=0}^{\infty} \beta^\tau \, (A^{o\prime})^\tau \;\, Z_a \, A^{o\tau} \tag{9}$$

$$\sigma_a = \frac{\beta}{1-\beta} \text{ trace } Z_a \sum_{\tau=0}^{\infty} \beta^\tau \, (A_o)^\tau \, CC'(A^{o\prime})^\tau . \tag{10}$$

According to (8), the asset price a_0 turns out to be the sum of a constant σ_a, which reflects a "risk premium," and a quadratic form in the state vector x_t. To understand why σ_a reflects a risk premium, notice that the parameters in C that govern the covariance matrix of innovations to the state influence σ_a but do not influence μ_a.

To implement (8) requires the application of numerical methods to calculate the matrices μ_a and σ_a that satisfy Eqs. (9) and (10). An efficient doubling algorithm for calculating these matrices is described below.

A nonlinear economy

An alternative method for parameterizing linear-quadratic economies is to generate them as approximations to non-linear-quadratic economies by using quadratic approximations to preferences and linear approximations to transition laws. These approximations make the parameters in the linear-quadratic structure functions of deeper parameters in the underlying economy.

Here is a version of Kydland and Prescott's (1982) method for using linear-quadratic control theory to compute approximate linear solutions to economies that are not linear-quadratic. Consider a social planning problem of the form

$$\max_{\{u_t\}} \; E_0 \sum_{t=0}^{\infty} \beta^t \, r(z_t, \theta)$$
$$\text{subject to } x_{t+1} = A x_t + B u_t + C w_{t+1}$$
$$z_t = [x'_t, u'_t]',$$

where θ is a vector of parameters and r is a function of the type used in the literatures on stochastic growth and real business cycles and w_t is a vector white noise.[3] Kydland and Prescott generate an approximate solution of this problem by solving a related problem:

$$\max_{\{u_t\}} E_0 \sum_{t=0}^{\infty} \beta^t \, z'_t M z_t$$
$$x_{t+1} = A x_t + B u_t + C w_{t+1},$$

[3] In most cases, r is the utility function after nonlinear constraints have been substituted in.

where

$$M = e(r(\bar{z},\theta) - \frac{\partial r(\bar{z},\theta)'}{\partial \bar{z}}\bar{z} + \frac{1}{2}\bar{z}'\frac{\partial^2 r(\bar{z},\theta)'}{\partial \bar{z}^2}\bar{z})e' + \frac{1}{2}(e\frac{\partial r(\bar{z},\theta)'}{\partial \bar{z}} + \frac{\partial r(\bar{z},\theta)}{\partial \bar{z}}e' - e\bar{z}'\frac{\partial^2 r(\bar{z},\theta)}{\partial \bar{z}^2} - \frac{\partial^2 r(\bar{z},\theta)}{\partial \bar{z}^2}ze' + \frac{\partial^2 r(\bar{z},\theta)}{\partial \bar{z}^2}),$$

where e is a vector of zeros with 1 in the element corresponding to the constant term in x_t, and $S_x = [I_n, 0_{n,k}]$ and $S_u = [0_{k,n}, I_k]$ are selector matrices and imply $z_t = S_x x_t + S_u u_t$, where n is the dimension of x_t and k is the dimension of u_t. This approximating problem is an optimal linear regulator problem.

Linear-Quadratic Models with Distortions

The computational procedures under study were originally applied to economies for which a competitive equilibrium allocation solves a social planning problem in the form of an optimal linear regulator problem and for which equilibrium prices (or approximations to them) can be deduced from the value function for the social planner. Most of the methods can, with some adaptations, also be used to study economies with particular types of externalities and other distortions, like taxes. Such adaptations are described by Blanchard and Kahn (1980), Whiteman (1983), Dagli and Taylor (1984), King, Plosser, and Rebelo (1988a,b), Hansen and Sargent (1994), and McGrattan (1994).

In linear-quadratic economies, the approach is to formulate the choice problem of a representative agent as a version of a linear regulator, while keeping account of the distinction between objects chosen by that agent, and economy-wide versions of those objects (the so-called "little k – big K" distinction, where the "little k" is chosen by the representative agent, taking "big K" as given, though in equilibrium "little k" = "big K"). The representative agent's problem is

$$\max_{\{\bar{u}_t\}} \sum_{t=0}^{\infty} \beta^t \{ \begin{bmatrix} \bar{y}_t \\ \bar{z}_t \end{bmatrix}' \begin{bmatrix} \bar{Q}_y & \bar{Q}_z \\ \bar{Q}_z' & \bar{Q}_{22} \end{bmatrix} \begin{bmatrix} \bar{y}_t \\ \bar{z}_t \end{bmatrix} + \bar{u}_t' \bar{R} \bar{u}_t + 2 \begin{bmatrix} \bar{y}_t \\ \bar{z}_t \end{bmatrix}' \begin{bmatrix} \bar{W}_y \\ \bar{W}_z \end{bmatrix} \bar{u}_t \}$$

subject to

$$\bar{y}_{t+1} = \bar{A}_y \bar{y}_t + \bar{A}_z \bar{z}_t + \bar{B}_y \bar{u}_t,$$

where $\bar{u}_t$ is a vector of controls set by the agent; $\bar{y}_t$ is a vector of state variables consisting of two types of variables, first, state variables under the partial control of the representative agent (the "little k" variables), and, second, stochastic processes like technology or preference shocks that are exogenous to the model; and $\bar{z}_t$ consists of a vector of state variables that are exogenous to the representative agent (the "big K" variables), but not to the model. The representative agent takes the sequence $\{z_t\}$ as given when solving this problem, even though after equilibrium is imposed the individual's choices determine the behavior of $\{z_t\}$.

In equilibrium (i.e., after the agent has optimized), the following equations must be satisfied:

$$\bar{z}_t = \bar{\Theta}\bar{y}_t + \bar{\Psi}\bar{u}_t.$$

Included in these equations would be the "big $K =$ little k" conditions.

Despite the fact the equilibrium allocation for this economy does not solve a social planning problem, it remains possible to compute an equilibrium by using algorithms closely related to ones that solve linear regulator problems. McGrattan (1994) gives details.

The Optimal Linear Regulator Problem

Consider the following version of the optimal linear regulator problem: choose a contingency plan for $\{u_t\}_{t=0}^{\infty}$ to maximize

$$E\sum_{t=0}^{\infty}\beta^t[x_t'Qx_t + u_t'Ru_t + 2x_t'Wu_t],\ 0<\beta<1 \tag{11}$$

subject to

$$x_{t+1} = Ax_t + Bu_t + Cw_{t+1},\ t\geq 0, \tag{12}$$

where x_0 is given. In (11) – (12), x_t is an $n\times 1$ vector of state variables and u_t is a $k\times 1$ vector of control variables. In (12), we assume that w_{t+1} is a martingale difference sequence with $Ew_tw_t' = I$ and that C is a matrix conformable as required to x and w.

We impose conditions on (Q,R,W) and (A,B) that are sufficient to imply that it is both feasible and desirable to set the controls in a way that implies that

$$E\sum_{t=0}^{\infty}\beta^t x_t'x_t \mid x_0 < \infty. \tag{13}$$

Dynamic programming

A standard way to solve this problem is by applying the method of dynamic programming. Let $V(x)$ be the optimal value associated with the program starting from initial state vector $x_0 = x$. Bellman's functional equation is

$$V(x_t) = \max_{u_t}\Big\{x_t'Qx_t + u_t'Ru_t + 2x_t'Wu_t + \beta E_tV(x_{t+1})\Big\}, \tag{14}$$

where the maximization is subject to (12). One way to solve this functional equation is simply to iterate on a version of Eq. (14), thereby constructing a sequence $V_j(x_t)$ of successively better approximations to $V(x_t)$. In particular, let

$$V_{j+1}(x_t) = \max_{u_t}\Big\{x_t'Qx_t + u_t'Ru_t + 2x_t'Wu_t + \beta E_tV_j(x_{t+1})\Big\}, \tag{15}$$

where again the maximization is subject to (12). Suppose that we initiate the iterations from $V_0(x)=0$. Then direct calculations show that successive iterations on Eq. (15) yield the quadratic form

$$V_j(x_t) = x_t' P_j x_t + \rho_j, \tag{16}$$

where P_j and ρ_j satisfy the equations

$$P_{j+1} = Q + \beta A' P_j A - (\beta A' P_j B + W)(R + \beta B' P_j B)^{-1}(\beta B' P_j A + W') \tag{17}$$
$$\rho_{j+1} = \beta \rho_j + \beta \text{ trace } P_j CC'. \tag{18}$$

Equation (17) is known as the *matrix Riccati difference equation.* Notice that it involves only $\{P_j\}$ and is independent of $\{\rho_j\}$. Notice also that the parameters in C, which multiplies the noises impinging on the system and so determines the variances of innovations to information in the system, affect the $\{\rho_j\}$ sequence but not the $\{P_j\}$ sequence. This fact can be summarized by saying that $\{P_j\}$ is independent of the system's noise statistics.

Under some regularity conditions described by Kwakernaak and Sivan (1972) and Sargent (1980), iterations on Eqs. (17) and (18) converge.[4] Let P and ρ be the limits of (17) and (18), respectively. Then the value function $V(x_t)$ that satisfies Bellman's equation (14) is given by

$$V(x_t) = x_t' P x_t + \rho,$$

where P and ρ are the limit points of iterations on (17) and (18) starting from $P_0 = 0, \rho_0 = 0$.

The decision rule that attains the right side of (15) is given by

$$u_t = -F_j x_t,$$

where

$$F_j = (R + \beta B' P_j B)^{-1}(\beta B' P_j A + W'). \tag{19}$$

The optimal decision rule for the original problem is given by $u_t = -F x_t$, where $F = \lim_{j\to\infty} F_j$, or

$$F = (R + \beta B' P B)^{-1}(\beta B' P A + W'). \tag{20}$$

According to Eq. (20), the optimum decision rule for u_t is independent of the parameters C and so also of the noise statistics.

The limit point P of iterations on (17) evidently satisfies

$$\begin{aligned} P = Q + \beta A' P A - (\beta A' P B + W) \\ \times (R + \beta B' P B)^{-1}(\beta B' P A + W'). \end{aligned}$$

[4] See Sargent (1980) for a discussion of these conditions.

This equation in P is called the *algebraic matrix Riccati equation.*

One standard way to solve an optimal linear regulator problem is simply to iterate directly on Eqs. (17) and (18). However, faster algorithms are available. These methods solve the algebraic matrix Riccati equation without iterating directly on (17). Before we describe some faster algorithms, we shall describe two useful transformations that permit simplification of some of the formulas presented above.

Two Useful Transformations

Removing cross-products between states and controls

It is often simpler to study problems without cross-products between states and controls. A simple transformation eliminates such cross-products. Consider a linear regulator problem with objective function

$$E\sum_{t=0}^{\infty}\beta^t\left\{[x_t'\ u_t^{*\prime}]\begin{bmatrix}Q^* & W\\ W' & R\end{bmatrix}\begin{bmatrix}x_t\\ u_t^*\end{bmatrix}\right\} \tag{21a}$$

that is to be maximized with respect to the transition law

$$x_{t+1}=A^*x_t+Bu_t^*+Cw_{t+1}. \tag{21b}$$

Define the transformed control u_t by

$$u_t=u_t^*+R^{-1}W'x_t. \tag{22}$$

Notice that

$$u_t'Ru_t=[\,x_t'\ u_t^{*\prime}\,]\begin{bmatrix}WR^{-1}W' & W\\ W' & R\end{bmatrix}\begin{bmatrix}x_t\\ u_t^*\end{bmatrix}.$$

It follows that

$$[\,x_t'\ u_t^{*\prime}\,]\begin{bmatrix}Q^* & W\\ W' & R\end{bmatrix}\begin{bmatrix}x_t\\ u_t^*\end{bmatrix}=x_t'Qx_t+u_t'Ru_t,$$

where $Q=Q^*-WR^{-1}W'$. Further, notice that the transition law (12) can be represented as

$$x_{t+1}=Ax_t+Bu_t+Cw_{t+1},$$

where $A=A^*-BR^{-1}W'$.

Collecting results, we find that the regulator problem (21) is equivalent to the following regulator problem without cross-products between states and controls: choose $\{u_t\}$ to maximize

$$E\sum_{t=0}^{\infty}\beta^t[x_t'Qx_t+u_t'Ru_t] \tag{23}$$

subject to

$$x_{t+1} = Ax_t + Bu_t + Cw_{t+1}, \tag{24}$$

where

$$\begin{aligned} Q &= Q^* - WR^{-1}W' \\ A &= A^* - BR^{-1}W'. \end{aligned} \tag{25}$$

It is often convenient to avail ourselves of the opportunity afforded by this transformation to focus on problems without cross-products between states and controls.

Eliminating discounting

Consider the following discounted optimal linear regulator problem: choose a contingency plan for $\{u_t\}$ to maximize

$$E\sum_{t=0}^{\infty} \beta^t \{x_t'Qx_t + u_t'Ru_t\}, 0 < \beta < 1 \tag{26}$$

subject to

$$x_{t+1} = Ax_t + Bu_t + C\xi_{t+1}, \tag{27}$$

where $\{\xi_{t+1}\}$ is a martingale difference sequence with $E\{\xi_t\xi_t'\} = \Omega_t$. Consider the transformed variables

$$\begin{aligned} \tilde{x}_t &= \beta^{\frac{t}{2}} x_t \\ \tilde{u}_t &= \beta^{\frac{t}{2}} u_t. \end{aligned} \tag{28}$$

In terms of the transformed variables, Eqs. (26) and (27) can be rewritten as

$$E\sum_{t=0}^{\infty} (\tilde{x}_t'Q\tilde{x}_t + \tilde{u}_t'R\tilde{u}_t) \tag{29}$$

$$\tilde{x}_{t+1} = \tilde{A}\tilde{x}_t + \tilde{B}\tilde{u}_t + C\beta^{\frac{t+1}{2}}\xi_{t+1}, \tag{30}$$

where

$$\begin{aligned} \tilde{A} &= \beta^{\frac{1}{2}} A \\ \tilde{B} &= \beta^{\frac{1}{2}} B \end{aligned} \tag{31}$$

and $E(\beta^{\frac{t+1}{2}}\xi_{t+1})(\beta^{\frac{t+1}{2}}\xi_{t+1})' = \beta^{t+1}\Omega_{t+1}$. The transformed optimal linear regulator problem is to choose a contingency plan for $\{\tilde{u}_t\}$ to maximize (29) subject to (30). The optimal control law for $\tilde{u}_t$ is given by

$$\tilde{u}_t = -\tilde{F}\tilde{x}_t,$$

where

$$\tilde{F} = (\tilde{B}'\tilde{P}\tilde{B} + R)^{-1}\tilde{B}'\tilde{P}\tilde{A}, \tag{32}$$

where $\tilde{P}$ is the limit point of iterations on an appropriate version of the matrix Riccati difference equation (17). The limit point $\tilde{P}$ thus satisfies

$$\tilde{P} = Q + \tilde{A}'\tilde{P}\tilde{A} - \tilde{A}'\tilde{P}\tilde{B}(R + \tilde{B}'\tilde{P}\tilde{B})^{-1}\tilde{B}'\tilde{P}\tilde{A}. \tag{33}$$

This is a version of the algebraic matrix Riccati equation. The optimal closed loop system in terms of transformed variables is

$$\tilde{x}_{t+1} = (\tilde{A} - \tilde{B}\tilde{F})\tilde{x}_t + \beta^{\frac{t+1}{2}} C\xi_{t+1}. \tag{34}$$

Multiplying both sides of this equation by $\beta^{-(\frac{t+1}{2})}$ gives

$$x_{t+1} = (A - B\tilde{F})x_t + C\xi_{t+1}. \tag{35}$$

Under standard assumptions on the undiscounted problem (29)-(30),[5] the eigenvalues of $(\tilde{A} - \tilde{B}\tilde{F})$ are less than unity in modulus. Since $A - B\tilde{F} = \beta^{-\frac{1}{2}}\ (\tilde{A} - \tilde{B}F)$, it follows that under these same assumptions about the *undiscounted* problem, the eigenvalues of $A - B\tilde{F}$ are less than $1/\sqrt{\beta}$ in modulus.

Vaughan's Eigenvector Method for Solving the Algebraic Matrix Riccati Equation

Vaughan (1970) described a fast algorithm for computing the limit point of the matrix Riccati equation (33). The multipliers in a Lagrangian formulation of the linear regulator problem can be represented in terms of derivatives of the value function. Vaughan's method works with the Lagrangian formulation of the problem and proceeds by deriving the linear restrictions that stability imposes across the multipliers and the state vector. Those restrictions can be used to compute the matrix P that solves the algebraic matrix Riccati equation.

Consider the following version of the optimal linear regulator problem: choose $\{u_t\}_{t=t_0}^{t_1-1}$ to maximize

$$\sum_{t=t_0}^{t_1-1}\{x_t'Qx_t + u_t'Ru_t\} + x_{t_1}'P_{t_1}x_{t_1} \tag{36}$$

subject to

$$x_{t+1} = Ax_t + Bu_t. \tag{37}$$

Let $\{\mu_t\}_{t=t_0+1}^{t_1}$ be a sequence of matrices of Lagrange multipliers. Form the Lagrangian

$$J = \sum_{t=t_0}^{t_1-1}\{x_t'Qx_t + u_t'Ru_t + 2\mu_{t+1}'[Ax_t + Bu_t - x_{t+1}]\} + x_{t_1}'P_{t_1}x_{t_1}. \tag{38}$$

[5] Again, see Sargent (1980).

First-order necessary conditions for the maximization of J with respect to $\{u_t\}_{t=t_0}^{t_1-1}$ and $\{x_t\}_{t=t_0}^{t_1-1}$ are

$$\begin{aligned} u_t: &\quad 2Ru_t + 2B'\mu_{t+1} = 0, &\quad t = t_0, \dots, t_1 - 1 &\quad (39)\\ x_t: &\quad \mu_t = Qx_t + A'\mu_{t+1}, &\quad t = t_0 + 1, \dots, t_1 - 1 &\quad (40)\\ &\quad \mu_t = P_{t_1} x_t, &\quad t = t_1. &\quad (41) \end{aligned}$$

Solve Eq. (39) for u_t and substitute into Eq. (37) to obtain

$$x_{t+1} = Ax_t - BR^{-1}B'\mu_{t+1}. \tag{42}$$

Stack Eqs. (39) and (40) to obtain

$$\begin{bmatrix} x_{t+1} \\ \mu_t \end{bmatrix} = \begin{bmatrix} A & -BR^{-1}B' \\ Q & A' \end{bmatrix} \begin{bmatrix} x_t \\ \mu_{t+1} \end{bmatrix}. \tag{43}$$

For the finite horizon problem, equation (43) is to be solved subject to the two boundary conditions, x_{t_0} given and $\mu_{t_1} = P_{t_1} x_{t_1}$.

To solve the infinite horizon problem that emerges when we set $t_1 = \infty$, Vaughan proceeded as follows. Assume that A is nonsingular. Then represent Eq. (43) as

$$\begin{bmatrix} x_t \\ \mu_t \end{bmatrix} = \begin{bmatrix} A^{-1} & A^{-1}BR^{-1}B' \\ QA^{-1} & QA^{-1}BR^{-1}B' + A' \end{bmatrix} \begin{bmatrix} x_{t+1} \\ \mu_{t+1} \end{bmatrix} \tag{44}$$

or

$$\begin{bmatrix} x_t \\ \mu_t \end{bmatrix} = M \begin{bmatrix} x_{t+1} \\ \mu_{t+1} \end{bmatrix}. \tag{45}$$

The matrix M is *symplectic*, which implies that its eigenvalues come in reciprocal pairs.[6] Assume that the eigenvalues of M are distinct, so that M has the representation

$$M = WDW^{-1}, \tag{46}$$

where D is a diagonal matrix of the eigenvalues of M, W is a matrix composed of the corresponding eigenvectors of M, and where D can be represented as

$$D = \begin{pmatrix} \Lambda & 0 \\ 0 & \Lambda^{-1} \end{pmatrix}, \tag{47}$$

where Λ is a diagonal matrix composed entirely of eigenvalues whose modulus exceeds unity. Because the eigenvalues appear in reciprocal pairs, we know that a representation of the form (46) – (47) exists for M.

Multiply both sides of (45) by M^{-1} to obtain

$$\begin{bmatrix} x_{t+1} \\ \mu_{t+1} \end{bmatrix} = W \begin{bmatrix} \Lambda^{-1} & 0 \\ 0 & \Lambda \end{bmatrix} \begin{bmatrix} V_{11}x_t + V_{12}\mu_t \\ V_{21}x_t + V_{22}\mu_t \end{bmatrix}. \tag{48}$$

[6] See Anderson and Moore (1979, p. 160) for a treatment of the key properties of symplectic matrices.

where $W^{-1} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$. Iterating on Eq. (48) j times gives

$$\begin{bmatrix} x_{t+j} \\ \mu_{t+j} \end{bmatrix} = W \begin{bmatrix} \Lambda^{-j} & 0 \\ 0 & \Lambda^{j} \end{bmatrix} \begin{bmatrix} V_{11}x_t & + & V_{12}\mu_t \\ V_{21}x_t & + & V_{22}\mu_t \end{bmatrix}, \tag{49}$$

where recall that the eigenvalues, the diagonal elements of Λ all exceed unity in modulus.

We want to solve Eq. (49) under conditions that imply that it is optimal to drive $x_t \to 0$ as $t \to \infty$, starting from any initial x_{t_0}. Since each component of Λ exceeds unity, the way to assure that $x_t \to 0$ as $t \to \infty$ is to insist that the components of the solution Eq. (49) multiplying Λ^j be set to zero. This is accomplished by setting the shadow prices μ_t to satisfy

$$\begin{aligned} & V_{21}x_t + V_{22}\mu_t = 0 \\ \text{or} \quad & \mu_t = -\, V_{22}^{-1} V_{21} x_t. \end{aligned} \tag{50}$$

Equation (50) states that μ_t is a particular time invariant linear function of x_t; call it $\mu_t = Px_t$, where $P = -V_{22}^{-1}V_{21}$. Under restriction (50), (49) becomes

$$\begin{bmatrix} x_{t+j} \\ \mu_{t+j} \end{bmatrix} = \begin{bmatrix} W_{11}\Lambda^{-j}(V_{11}x_t + V_{12}\mu_t) \\ W_{21}\Lambda^{-j}(V_{11}x_t + V_{12}\mu_t) \end{bmatrix}. \tag{51}$$

However, we know that $\mu_t = Px_t$. Therefore, Eq. (51) implies that

$$\begin{bmatrix} Px_{t+j} \\ \mu_{t+j} \end{bmatrix} = \begin{bmatrix} PW_{11}\Lambda^{-j}\ (V_{11}x_t + V_{12}\mu_t) \\ W_{21}\Lambda^{-j}\ (V_{11}x_t + V_{12}\mu_t) \end{bmatrix},$$

which implies that $PW_{11} = W_{21}$ or

$$P = W_{21}W_{11}^{-1}. \tag{52}$$

Equation (52) is Vaughan's equation for the solution of the algebraic matrix Riccati equation.

An Algorithm for Distorted Systems

Vaughan's method can be adapted to compute equilibria of models whose allocations do not solve a dynamic programming problem. Consider the problem: choose $\{u_t\}_{t=t_0}^{t_1-1}$ to

$$\max_{\{u_t\}} \sum_{t=t_0}^{t_1-1} \left\{ \begin{bmatrix} y_t \\ z_t \end{bmatrix}' \begin{bmatrix} Q_y & Q_z \\ Q_z' & Q_{22} \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix}' + u_t' R u_t \right\} \tag{53}$$

subject to

$$y_{t+1} = A_y y_t + A_z z_t + B_y u_t. \tag{54}$$

We have used the tricks described earlier to convert our original problem to one without discounting or cross-products between states and controls. In equilibrium, we assume that the following conditions must also be satisfied:

$$z_t = \Theta y_t + \Psi u_t. \tag{55}$$

First-order necessary conditions with respect to $\{u_t\}_{t=t_0}^{t_1-1}$ and $\{y_t\}_{t=t_0}^{t_1-1}$ in this case are given by

$$\begin{aligned}
u_t :& \quad 2Ru_t + 2B_y'\mu_{t+1} = 0, && t = t_0, \ldots, t_1 - 1 && (56)\\
y_t :& \quad \mu_t = Q_y y_t + Q_z z_t + A_y'\mu_{t+1}, && t = t_0 + 1, \ldots, t_1 - 1 && (57)\\
& \quad \mu_t = P_{t_1}[y_t', z_t'], && t = t_1, && (58)
\end{aligned}$$

where $\{\mu_t\}$ are Lagrange multipliers associated with the constraint in Eq. (54). Solve Eq. (56) for u_t and substitute it and Eq. (55) into Eqs. (54) and (57) to obtain

$$y_{t+1} = (A_y + A_z\Theta)y_t - (B_y + A_z\Psi)R^{-1}B_y'\mu_{t+1}, \tag{59}$$

$$\mu_t = (Q_y + Q_z\Theta)y_t + (A_y' - Q_z\Psi R^{-1}B_y')\mu_{t+1}. \tag{60}$$

Note that this system is similar to that of (43) in the undistorted case. To solve the infinite horizon problem that emerges when we set $t_1 = \infty$, proceed as follows. Assume that the matrix $A_y + A_z\Theta$ is nonsingular.[7] Then represent Eqs. (59) and (60) as

$$\begin{bmatrix} y_t \\ \mu_t \end{bmatrix} = \begin{bmatrix} \hat{A}^{-1} & \hat{A}^{-1}\hat{B}R^{-1}B_y' \\ \hat{Q}\hat{A}^{-1} & \hat{Q}\hat{A}^{-1}\hat{B}R^{-1}B_y' + \tilde{A}' \end{bmatrix} \begin{bmatrix} y_{t+1} \\ \mu_{t+1} \end{bmatrix} \tag{61}$$

or

$$\begin{bmatrix} y_t \\ \mu_t \end{bmatrix} = M \begin{bmatrix} y_{t+1} \\ \mu_{t+1} \end{bmatrix}, \tag{62}$$

where $\hat{A} = A_y + A_z\Theta$, $\hat{Q} = Q_y + Q_z\Theta$, $\hat{B} = B_y + A_z\Psi$, and $\tilde{A} = A_y - B_yR^{-1}\Psi'Q_z'$. Notice that if we replace $\hat{A}$ and $\tilde{A}$ with A, $\hat{B}$ and B_y with B, and $\hat{Q}$ with Q, then we have the same system as in (44). The differences between the systems occur because of the side conditions in Eq. (55) that must be satisfied. Notice also that in the case with distortions, M is not necessarily symplectic. We assume, however, that M has a representation

$$M = WDW^{-1}, \tag{63}$$

[7] See McGrattan (1994) for details of the finite horizon case and cases in which $\hat{A}$ is singular.

where D is a diagonal matrix of the eigenvalues of M, W is a matrix composed of the corresponding eigenvectors of M, and D can be represented as

$$D = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix}, \tag{64}$$

where Λ_1 is a diagonal matrix composed entirely of eigenvalues whose modulus exceeds unity, Λ_2 is a diagonal matrix composed entirely of eigenvalues whose modulus is below unity, and the dimensions of Λ_1 and Λ_2 are equal. We assume that Λ_1 and Λ_2 have equal numbers of eigenvalues, a condition for there to exist a unique bounded solution. In practice, we would check this condition during the calculations.

From this point on, we can follow the same procedure as in the previous section. Partition W, i.e.,

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}, \tag{65}$$

into four subpartitions of equal dimension. Set $\mu_t = W_{21}W_{11}^{-1}y_t$ so that $y_t \to 0$ as $t \to \infty$. Substitute this expression for μ_t into Eq. (59) to get μ_{t+1} in terms of y_t, i.e.,

$$\mu_{t+1} = (P^{-1} + \hat{B}R^{-1}B_y')^{-1}\hat{A}y_t, \tag{66}$$

where $P = W_{21}W_{11}^{-1}$. Therefore, the solution to the problem in Eq. (53) is given by

$$u_t = -R^{-1}B_y'(P^{-1} + \hat{B}R^{-1}B_y')^{-1}\hat{A}y_t. \tag{67}$$

Note that if $\Theta = 0$ and $\Psi = 0$, then Eq. (67) is identical to the optimal decision rule for the social planner of an undistorted economy linear-quadratic economy.

A Doubling Algorithm

To compute asset prices and to solve a Riccati equation using the partitioning methods described below, we have cause to compute infinite sums of the form

$$V = \sum_{j=0}^{\infty} G'^j D H^j,$$

where the eigenvalues of G and H are bounded in modulus strictly below unity. This sum can be evaluated by recognizing that it is the solution of a discrete Lyapunov equation and using an algorithm to solve that kind of equation. Alternatively, it could be computed by iterating to convergence on

$$V_{j+1} = D + G'V_jH.$$

Instead of using one of these methods, we often use a simple *doubling* algorithm, which we implement by computing the following objects recursively:

$$\begin{aligned} G_j &= G_{j-1}G_{j-1} \\ H_j &= H_{j-1}H_{j-1} \\ V_j &= V_{j-1} + G'_{j-1}V_{j-1}H_{j-1} \end{aligned} \tag{68}$$

where we set $V_0 = D,\ G_0 = G,\ H_0 = H$. By repeated substitution it can be shown that

$$V_j = \sum_{i=0}^{2^j-1} G^{i\prime} D H^i. \tag{69}$$

Each iteration doubles the number of terms in the sum.

The idea of accelerating convergence by skipping steps via doubling can be used to solve a Riccati equation.

Another Doubling Algorithm

The algebraic matrix Riccati equation can be solved by using a *doubling algorithm*.[8] The algorithm is related to Vaughan's method in the prominent role it assigns to the matrix M in Eq. (45).

We consider the same version of the optimal linear regulator focused on in Vaughan's method, namely, an undiscounted, nonstochastic problem without cross-products between states and controls. The problem is to choose a plan for $\{u_t\}_{t=t_0}^{t_1-1}$ to maximize

$$\sum_{t=t_0}^{t_1-1}\{x_t'Qx_t + u_t'Ru_t\} + x_{t_1}'P_{t_1}x_{t_1} \tag{70}$$

subject to

$$x_{t+1} = Ax_t + Bu_t. \tag{71}$$

Let the value function for the tail of the problem starting from initial condition x_t at time t be $x_t'P_tx_t$, for $t = t_0,\ t_0+1,\ \ldots, t_1-1$. The matrix Riccati difference equation is

$$P_t = Q + A'P_{t+1}A - A'P_{t+1}B(R + B'P_{t+1}B)^{-1}B'P_{t+1}A. \tag{72}$$

The first step in deriving the doubling algorithm is to use some facts from linear algebra to show that Eq. (72) implies the following difference equation for P_t:

$$\begin{aligned} P_t = &\{QA^{-1} + [A' + QA^{-1}BR^{-1}B']P_{t+1}\} \\ &\times\{A^{-1} + A^{-1}BR^{-1}B'P_{t+1}\}^{-1}. \end{aligned} \tag{73}$$

[8] This section is based on Anderson and Moore (1979, pp. 158–160).

Equation (73) is of the form

$$P_t = \{C + DP_{t+1}\} \times \{E + FP_{t+1}\}^{-1}, \tag{74}$$

where

$$\begin{aligned} C &= QA^{-1} \\ D &= A' + QA^{-1}BR^{-1}B' \\ E &= A^{-1} \\ F &= A^{-1}BR^{-1}B'. \end{aligned} \tag{75}$$

We can represent the evolution of Eq. (75) via the equivalent system

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} E & F \\ C & D \end{bmatrix} \begin{bmatrix} X_{t+1} \\ Y_{t+1} \end{bmatrix}, \tag{76}$$

where $P_{t+1} = Y_{t+1}X_{t+1}^{-1}$ and $P_t = Y_t X_t^{-1}$. Notice that

$$\begin{bmatrix} E & F \\ C & D \end{bmatrix} = \begin{bmatrix} A^{-1} & A^{-1}BR^{-1}B' \\ QA^{-1} & A' + QA^{-1}BR^{-1}B' \end{bmatrix} \equiv M. \tag{77}$$

The matrix on the right side of Eq. (77) is the matrix M on the right side of Eq. (44) or (45). The solution of Eq. (76) can be computed rapidly by using the fact that the matrix M on the right side is a *symplectic matrix* and by exploiting the properties of symplectic matrices.

A symplectic matrix Z can be represented in the form

$$Z = \begin{bmatrix} \alpha^{-1} & \alpha^{-1}\beta \\ \gamma\alpha^{-1} & \alpha' + \gamma\alpha^{-1}\beta \end{bmatrix}. \tag{78}$$

Notice how the matrix in (77) is in such a form, where we set $\alpha = A$, $\gamma = Q$, $\beta = BR^{-1}B'$.

Represent Eq. (76) in the form

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = M \begin{bmatrix} X_{t+1} \\ Y_{t+1} \end{bmatrix}. \tag{79}$$

Take the eigenvector decomposition of M given in (46), namely,

$$M = W \begin{pmatrix} \Lambda & 0 \\ 0 & \Lambda^{-1} \end{pmatrix} W^{-1},$$

where the Λ is a diagonal matrix consisting of the eigenvalues of M that exceed unity in modulus. Represent M in the partitioned form

$$M = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & \Lambda^{-1} \end{bmatrix} \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix},$$

where $[V_{ij}]$ is the partition of W^{-1}. Iterating on the partitioned form of Eq. (79) k times and noting that the elements of Λ exceed unity in modulus, it follows that

$$\lim_{k\to\infty} P_{t-k+1} = \lim_{k\to\infty} Y_{t-k+1}X_{t-k+1}^{-1} = W_{21}W_{11}^{-1}, \tag{80}$$

which is a version of Vaughan's (1969) formula (51) for computing the solution of the algebraic matrix Riccati equation. In Eq. (80), we have established that for any terminal matrices X_{t+1}, Y_{t+1} that satisfy $P_{t+1} = Y_{t+1}X_{t+1}^{-1}$, the limit of $P_{t+1-k} = Y_{t+1-k}X_{t+1-k}^{-1}$ is the solution P, which determines the value function for the infinite horizon version of the optimal linear regulator problem.

To compute $\lim_{k\to\infty} P_{t-k+1}$ we can proceed by computing higher and higher powers of M. Rather than computing the sequence M, M^2, M^3, $\ldots$, the *doubling algorithm* proceeds by skipping steps and only computing the sequence $M, M^2, M^4, M^8, \ldots$. Define $\phi(1) = M$ and define $\phi(2) = M^2$ $= \phi(1)^2$. Then define

$$\phi(2^k) = \phi(2^{k-1})^2 \tag{81}$$

for $k = 2, 3 \ldots$. Evidently, we have that

$$\phi(2^k) = M^{2^k}, k = 1, 2, 3, \ldots.$$

Thus, we recursively compute the sequence M, M^2, M^4, M^8, $\ldots$, $M^{2^k}, \ldots$ by simply squaring the preceding element of the sequence. We represent the solution of (79) in the form

$$\begin{bmatrix} X_{t-2^k+1} \\ Y_{t-2^k+1} \end{bmatrix} = M^{2^k} \begin{bmatrix} X_{t+1} \\ Y_{t+1} \end{bmatrix}. \tag{82}$$

We can compute $P_{t-2^k+1} = Y_{t-2^k+1}X_{t-2^k+1}^{-1}$.

Equation (82) is the key to the doubling algorithm. The algorithm is completed with the following two details. First, one computes the squares of the matrices M^{2^k} by using the following algorithm for squaring symplectic matrices:

$$\begin{aligned} \alpha_{k+1} &= \alpha_k(I + \beta_k\gamma_k)^{-1}\alpha_k \\ \beta_{k+1} &= \beta_k + \alpha_k(I + \beta_k\gamma_k)^{-1}\beta_k\alpha_k' \\ \gamma_{k+1} &= \gamma_k + \alpha_k'\gamma_k(I + \beta_k\gamma_k)^{-1}\alpha_k, \end{aligned} \tag{83}$$

where we set $\alpha_0 = A$, $\gamma_0 = Q$, $\beta_0 = BR^{-1}B'$. With this algorithm, we have that

$$M^{2^k} = \begin{bmatrix} \alpha_k^{-1} & \alpha_k^{-1}\beta_k \\ \gamma_k\alpha_k^{-1} & \alpha_k' + \gamma_k\alpha_k^{-1}\beta_k \end{bmatrix}. \tag{84}$$

Second, with M^{2^k} given by Eq. (84) in Eq. (82), and setting $X_{t+1} = I$, $Y_{t+1} = 0$, we obtain

$$Y_{t-2^k+1}X_{t-2^k+1}^{-1} = \gamma_k. \tag{85}$$

Equality (85) implies that we can compute the solution P of the algebraic matrix Riccati equation from

$$P = \lim_{k \to \infty} \gamma_k, \tag{86}$$

where γ_k is computed via Eq. (83).

Even though it was assumed that A^{-1} exists in deriving the doubling algorithm, notice that in (83) there is no call to invert A. Indeed, the algorithm seems to work well even when A^{-1} does not exist.

It is worth noting that while γ_k converges as $k \to \infty$, neither α_k nor β_k converges. On the contrary, both α_k and β_k diverge at a rate determined by the eigenvalue in Λ that is largest in absolute value. The matrix M^{2^k} diverges as $k \to \infty$; what converges is the "ratio" $Y_{t-2^k} X_{t-2^k}^{-1}$.

The doubling algorithm is much faster than iterating on the Riccati equation because it skips so many steps.

Adding Speed by Partitioning the State Vector

After application of the two transformations described above to remove discounting and cross-products between states and controls, often our control problem occurs in a *controllability canonical form*: choose $\{u_t\}$ to maximize

$$\sum_{t=0}^{\infty} \{ \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}' \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} + u_t' R u_t \} \tag{87}$$

subject to

$$\begin{bmatrix} x_{1t+1} \\ x_{2t+1} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u_t, \tag{88}$$

with $[x_{10}', x_{20}']'$ given. The pattern of zeros in the partitioned versions of A and B in Eq. (88) reflect that x_{2t} is an "uncontrollable process" from the viewpoint of a social planner.[9] Two things distinguish a controllability canonical form: (1) the pattern of zeros in the pair (A, B) and (2) a requirement that (A_{11}, B_1) be a controllable pair, by which is meant that the matrix $[B_1 \ A_{11} B_1 \ A_{11}^2 B_1 \cdots A_{11}^{n-1} B_1]$ have rank equal to the dimension of A_{11}. A controllability canonical form adopts a description of the state vector that separates it into a part x_{2t} that cannot be affected by the controls and a part x_{1t} that can be controlled in the sense that there exists a sequence of controls $\{u_t\}$ that sends x_1 to any arbitrarily specified point within the space in which x_1 lives.

An advantage in working with a system in controllability canonical form is that computing the optimal controls can be simplified by organizing the calculations in a recursive way, first focusing on the controllable point of the system.

[9] See Kwakernaak and Sivan (1972) or Sargent (1980).

Define an operator T associated with Bellman's equation:

$$T(P) = Q + A'PA - A'PB(R + B'PB)^{-1}B'PA. \tag{89}$$

Partition P and $T(P)$ conformably with the partition $\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}$. The $(1,1)$ and $(1,2)$ components of $T(P)$ satisfy

$$T_{11}(P_{11}) = Q_{11} + A'_{11}P_{11}A_{11} - A'_{11}P_{11}B_1(R + B'_1P_{11}B_1)^{-1}B'_1P_{11}A_{11} \tag{90}$$

and

$$\begin{aligned} T_{12}(P_{11},P_{12}) = {} & Q_{12} + A'_{11}P_{11}A_{12} \\ & - A'_{11}P_{11}B_1(R + B'_1P_{11}B_1)^{-1}B'_1P_{11}A_{12} \\ & + [A'_{11} - A'_{11}P_{11}B_1(R + B'_1P_{11}B_1)^{-1}B'_1]P_{12}A_{22}. \end{aligned} \tag{91}$$

Notice from Eq. (90) that T_{11} depends only on P_{11} and not on other elements of the partition of P. From Eq. (91), T_{12} depends on P_{11} and P_{12}, but not on P_{22}. Because T maps symmetric matrices into symmetric matrices, the $(2,1)$ block of T is just the transpose of the $(1,2)$ block. Finally, the $(2,2)$ block of T depends on P_{11}, P_{12}, and P_{22}.

Partition the optimal control state feedback matrix $F = [F_1\ F_2]$, where the partition is conformable with that of x_t. The optimal control is

$$u_t = -[F_1\ F_2]\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}.$$

Let P^f_{11} be the fixed point of Eq. (90), and let P^f_{12} be the fixed point of $T_{12}(P^f_{11}, P_{12})$. Then F_1 and F_2 are given by

$$F_1 = (R + B'_1P^f_{11}B_1)^{-1}B'_1P^f_{11}A_{11} \tag{92}$$

$$F_2 = (R + B'_1P^f_{11}B_1)^{-1}(B'_1P^f_{11}A_{12} + B'_1P^f_{12}A_{22}). \tag{93}$$

Equation (92) shows that F_1 depends only on P^f_{11}, while F_2 depends on P^f_{11} and P^f_{12}, but not on P^f_{22}, the fixed point of T_{22}.

We can compute the fixed points of T_{11} and T_{12} as follows. First, note that the T_{11} operator identified by (90) is formally equivalent with the T operator of (89), except that $(1,1)$ subscripts appear on A and Q, and a (1) subscript appears on B. Thus, the T_{11} operator is simply the operator whose iterations define the matrix Riccati difference equation for the small optimal regulator problem determined by the matrixes $(A_{11},\ B_1,\ R,\ Q_{11})$. We can compute P^f_{11} by using any of the algorithms described above for this smaller problem.

Second, given a fixed point P^f_{11} of T_{11}, we apply our simple doubling algorithm to compute the fixed point of $T_{12}(P^f_{11}, \cdot)$. From (68), this mapping has the form

$$T_{12}(P^f_{11}, P_{12}) = D + G'P_{12}H, \tag{94}$$

where

$$
\begin{aligned}
D &= Q_{12} + A'_{11} P^f_{11} A_{12} - A'_{11} P^f_{11} B_1 (R + B'_1 P^f_{11} B_1)^{-1} B'_1 P^f_{11} A_{12} \\
G &= [A_{11} - B_1 (R + B'_1 P^f_{11} B_1)^{-1} B'_1 P^f_{11} A_{11}] \\
H &= A_{22}.
\end{aligned}
$$

Notice that $G = A_{11} - B_1 F_1$, where F_1 is computed from (92). When x_{2t} is set to zero for all t, the law of motion for x_{1t} under the optimal control is thus given by

$$x_{1t+1} = G x_{1t}.$$

We have assumed regularity conditions that are sufficient to imply that the eigenvalues of G have absolute values strictly less than unity. The eigenvalues of H also are strictly less than unity by assumption. That the eigenvalues of G and H are both less than unity assures the existence of a limit point to iterations on Eq. (94). The limit point of iterations on $T_{12}(P^f_{11}, P_{12})$ starting from $P_{12} = 0$ can be represented as

$$P^f_{12} = \sum_{j=0}^{\infty} G'^j D H^j. \tag{95}$$

We compute P^f_{12} by using the doubling algorithm described above.

Innovations Representations

Constructing an *innovations representation* is a key step in deducing the implications of a model for vector autoregressions and for evaluating a Gaussian likelihood function.[10] An innovations representation is a state-space representation in which the vector white noise driving the system is of the correct dimension (equal to that of the vector of observables) and lives in the proper space (the space spanned by current and lagged values of the observables).

Suppose that our theorizing and data collection lead us to a system of the form

$$
\begin{aligned}
x_{t+1} &= A_o x_t + C w_{t+1} \\
z_t &= G x_t + v_t \\
v_t &= D v_{t-1} + \eta_t,
\end{aligned}
\tag{96}
$$

where D is a matrix whose eigenvalues are bounded in modulus by unity and η_t is a martingale difference sequence that satisfies

$$
\begin{aligned}
E \eta_t \eta'_t &= R \\
E w_{t+1} \eta'_s &= 0 \quad \text{for all } t \text{ and } s.
\end{aligned}
$$

[10] The calculations in this section are versions of ones described by Anderson and Moore (1979).

In Eq. (96), v_t is a serially correlated measurement error process that is orthogonal to the x_t process.

We define the quasi-differenced process as

$$\bar{z}_t \equiv z_{t+1} - Dz_t. \tag{97}$$

From Eq. (96) and the definition (97) it follows that

$$\bar{z}_t = (GA_o - DG)x_t + GCw_{t+1} + \eta_{t+1}.$$

Then $(x_t, \bar{z}_t)$ is governed by the state-space system

$$\begin{aligned} x_{t+1} &= A_o x_t + Cw_{t+1} \\ \bar{z}_t &= \bar{G}x_t + GCw_{t+1} + \eta_{t+1}, \end{aligned} \tag{98}$$

where $\bar{G} = GA_o - DG$. This system has nonzero covariance between the state noise Cw_{t+1} and the "measurement noise" $(GCw_{t+1} + \eta_{t+1})$. Let $[K_t, \Sigma_t]$ be the Kalman gain and state covariance matrix associated with the Kalman filter, namely,

$$K_t = (CC'G' + A_o\Sigma_t\bar{G}')\Omega_t^{-1} \tag{99}$$

$$\Omega_t = \bar{G}\Sigma_t\bar{G}' + R + GCC'G' \tag{100}$$

$$\Sigma_{t+1} = A_o\Sigma_t A_o{}' + CC' - (CC'G' + A_o\Sigma_t\bar{G}')\Omega_t^{-1}(\bar{G}\Sigma_t A_o{}' + GCC'). \tag{101}$$

Then an innovations representation for system (98) is

$$\begin{aligned} \hat{x}_{t+1} &= A_o\hat{x}_t + K_t u_t \\ \bar{z}_t &= \bar{G}\hat{x}_t + u_t, \end{aligned} \tag{102}$$

where

$$\begin{aligned} \hat{x}_t &= \hat{E}[x_t \mid \bar{z}_{t-1}, \bar{z}_{t-2}, \ldots, \bar{z}_0, \hat{x}_0] \\ u_t &= \bar{z}_t - \hat{E}[\bar{z}_t \mid \bar{z}_{t-1}, \ldots, \bar{z}_0, \hat{x}_0] \\ \Omega_t &\equiv Eu_t u_t' = \bar{G}\Sigma_t\bar{G}' + R + GCC'G'. \end{aligned}$$

Initial conditions for the system are $\hat{x}_0$ and Σ_0. From definition (97), it follows that $[z_{t+1}, z_t, \ldots, z_0, \hat{x}_0]$ and $[\bar{z}_t, \bar{z}_{t-1}, \ldots, \bar{z}_0, \hat{x}_0]$ span the same space, so that

$$\begin{aligned} \hat{x}_t &= \hat{E}[x_t \mid z_t, z_{t-1}, \ldots, z_0, \hat{x}_0] \\ u_t &= z_{t+1} - \hat{E}[z_{t+1} \mid z_t, \ldots, z_0, \hat{x}_0]. \end{aligned}$$

So u_t is said to be an innovation in z_{t+1}.

Equation (101) is a matrix Riccati difference equation. The Kalman filter has a steady-state solution if there exists a time-invariant matrix Σ which satisfies Eq. (101), i.e., one that satisfies the algebraic matrix Riccati equation. In this case, the same computational procedures used for the optimal linear regulator problem apply. This is a benefit of the duality of filtering and control referred to earlier. The steady-state Kalman gain, K, is given by Eq. (99) with $\Sigma_t = \Sigma$ and $\Omega_t = \bar{G}\Sigma\bar{G}' + R + GCC'G'$.

The innovations representation is equivalent with a *Wold representation* or *vector autoregression*. Estimates of these representations are recovered in empirical work using the vector autoregressive techniques promoted by Sims (1980) and Doan, Litterman, and Sims (1984). It is convenient to have a quick way of deducing the vector autoregression implied by a particular theoretical structure. To get a Wold representation for z_t, substitute Eq. (97) into Eq. (102) to obtain

$$\begin{aligned} \hat{x}_{t+1} &= A_o\hat{x}_t + Ku_t \\ z_{t+1} - Dz_t &= \bar{G}\hat{x}_t + u_t. \end{aligned} \tag{103}$$

A Wold representation for z_t is

$$z_{t+1} = [I - DL]^{-1}[I + \bar{G}(I - A_oL)^{-1}KL]u_t, \tag{104}$$

where again L is the lag operator. From Eq. (103) a recursive whitening filter for obtaining $\{u_t\}$ from $\{z_t\}$ is given by

$$\begin{aligned} u_t &= z_{t+1} - Dz_t - \bar{G}\hat{x}_t \\ \hat{x}_{t+1} &= A_o\hat{x}_t + Ku_t. \end{aligned} \tag{105}$$

Vector autoregressive representation

Hansen and Sargent (1994) show that an autoregressive representation for z_t is

$$z_{t+1} = \{D + (I - DL)\,\bar{G}[I - (A_o - K\bar{G})L]^{-1}KL\}\,z_t + u_t. \tag{106}$$

or

$$\begin{aligned} z_{t+1} =& [D + \bar{G}K]z_t + \sum_{j=1}^{\infty}[\bar{G}(A_o - K\bar{G})^jK \\ &- D\bar{G}(A_o - K\bar{G})^{j-1}K]z_{t-j} + u_t. \end{aligned} \tag{107}$$

This equation expresses z_{t+1} as the sum of the one-step-ahead linear least squares forecast and the one-step prediction error.

The Likelihood Function

We start with a "raw" time series $\{y_t\}$ that determines an adjusted series z_t according to

$$z_t = f(y_t, \Theta),$$

where Θ is the vector containing the free parameters of the model, including parameters determining particular detrending procedures. For example, if our raw series has a geometric growth trend equal to μ^t which is to be removed before estimation, then the adjusted series is $z_t = y_t/\mu^t$. We assume that the state-space model of the form (98) and the associated innovations representation (102) pertains to the adjusted data $\{z_t\}$. We can use the innovations representation (102) recursively to compute the innovation series, then calculate the log-likelihood function

$$L(\Theta) = \sum_{t=0}^{T-1} \{ \log |\Omega_t| + \text{trace}(\Omega_t^{-1} u_t u_t') - \log |\frac{\partial f(y_t, \Theta)}{\partial y_t}| \} \tag{108}$$

and find estimates, $\hat{\Theta} = \text{argmin}_\Theta L(\Theta)$, where $\Omega_t = E u_t u_t'$ is the covariance matrix of the innovations. To find the minimizer $\hat{\Theta}$, we can use a standard optimization program. In practice, it is best if we can calculate both the log-likelihood function and its derivatives analytically. First, the computational burden is much lower with analytical derivatives. Consider, for example, the model of McGrattan, Rogerson, and Wright (1993), which has 84 elements in Θ. For each step of a quasi-Newton optimization routine, L and $\frac{\partial L}{\partial \theta}$ are computed. To obtain $\frac{\partial L}{\partial \theta}$ numerically for the McGrattan, Rogerson, Wright (1993) example, the log-likelihood function must be evaluated 168 times if central differences are used in computing an approximation for $\frac{\partial L}{\partial \theta}$, e.g.,

$$\frac{\partial L}{\partial \theta} \approx \frac{L(\Theta + \epsilon e) - L(\Theta - \epsilon e)}{2\epsilon}, \tag{109}$$

where e is a vector of zeros except for a 1 in the element corresponding to θ and ϵ is some positive number. Usually, the costs of computing L a large number of times far outweigh the costs of computing $\frac{\partial L}{\partial \theta}$ once. If L and $\frac{\partial L}{\partial \theta}$ are to be computed many times, which is typically the case, then the costs of computing numerical derivatives can be quite large. A second advantage to analytical derivatives is numerical accuracy. If the log-likelihood function is not very smooth for the entire parameter space, there may be problems with the accuracy of approximations such as Eq. (109). With inaccurate derivatives, it is difficult to determine the curvature of the function and, hence, to find a minimum.

For $L(\Theta)$ in Eq. (108), the derivatives $\partial L(\Theta)/\partial \theta$ are easy to derive. We derive them in Appendix A and distinguish formulas that are steps in the derivation from those that would be put into a computer code. Note that although the final expression for $\frac{\partial L}{\partial \theta}$ derived in Appendix A is complicated, we can use numerical approximations such as Eq. (109) to uncover coding errors.

Once we have the log-likelihood function and its derivatives, we can apply standard optimization methods to the problem of finding the maximum likelihood estimates. In practice, we will have a constrained optimization problem since the equilibrium is not typically computable for all possible parameterizations. For example, we may have simple constraints such as $\ell < \Theta < u$, where ℓ and u are the lower and upper bounds for the parameter vector.

In this case, we use either a constrained optimization package or penalty functions (see Fletcher 1987).

After computing the maximum likelihood estimates, we need to compute their standard errors,

$$S_e(\Theta) = \operatorname{diag}\left(\sqrt{(\sum_t \frac{\partial L_t}{\partial \Theta}\frac{\partial L_t}{\partial \Theta}'})^{-1}\right), \tag{110}$$

where $L_t(\Theta)$ is the logarithm of the density function of the date t innovation, i.e.,

$$L_t(\Theta) = \log|\Omega_t| + u_t'\Omega_t^{-1}u_t - \log\left|\frac{\partial f(y_t,\Theta)}{\partial y_t}\right|. \tag{111}$$

The formula for $\frac{\partial L_t}{\partial \theta}$ is also given in Appendix A.

An Example

In this section, we present estimates of some of the parameters of Rosen, Murphy, and Scheinkman's (1994) model of "Cattle Cycles." Let p_t be the price of freshly slaughtered beef, m_t the feeding cost of preparing an animal for slaughter, h_t the one-period holding cost for a mature animal, $\gamma_1 h_t$ the one-period holding cost for a yearling, and $\gamma_0 h_t$ the one-period holding cost for a calf. The costs $\{h_t, m_t\}_{t=0}^{\infty}$ are exogenous stochastic processes, while the stochastic process $\{p_t\}_{t=0}^{\infty}$ is determined by a rational expectations equilibrium. Let x_t be the breeding stock and y_t be the total stock of animals. Each animal that is reserved for breeding gives birth to g calves. Calves that survive become part of the adult stock after 2 years. Therefore, if we assume that t indexes a year, the law of motion for stocks is

$$x_t = (1-\delta)x_{t-1} + gx_{t-3} - c_t, \tag{112}$$

where c_t is a rate of slaughtering and δ is the exponential death rate. The total head count of cattle is

$$y_t = x_t + gx_{t-1} + gx_{t-2}, \tag{113}$$

which is the sum of adults, yearlings, and calves, respectively.

A representative farmer maximizes

$$\begin{aligned} E_0 \sum_{t=0}^{\infty} \beta^t \{ p_t c_t - h_t x_t - (\gamma_0 h_t)(gx_{t-1}) - (\gamma_1 h_t)(gx_{t-2}) - m_t c_t \\ - \Psi(x_t, x_{t-1}, x_{t-2}, c_t)\}, \end{aligned} \tag{114}$$

where

$$\Psi = \frac{\psi_1}{2}x_t^2 + \frac{\psi_2}{2}x_{t-1}^2 + \frac{\psi_3}{2}x_{t-2}^2 + \frac{\psi_4}{2}c_t^2. \tag{115}$$

The maximization is subject to the law of motion (112), taking as given the stochastic laws of motion for the exogenous random processes and the equilibrium price process and the initial state $[x_{-1},\ x_{-2},\ x_{-3}]$. Here $(\psi_j, j = 1, 2, 3)$ are small positive parameters, which model quadratic costs of carrying stocks, and ψ_4 is a small positive parameter measuring quadratic costs of slaughtering.[11]

Demand is governed by

$$c_t = \alpha_0 - \alpha_1 p_t + d_t, \tag{116}$$

where $\alpha_0 > 0$, $\alpha_1 > 0$, and $\{d_t\}_{t=0}^{\infty}$ is a stochastic process with mean zero representing a demand shifter. The stochastic processes $\{d_t,\ b_t,\ m_t\}$ are univariate autoregressions with orthogonal innovations.

We can map this model into the framework of Hansen and Sargent (1994) by appropriately choosing the matrices Λ, Π, Θ_h, Δ_h, Δ_k, Θ_k, Φ_c, Φ_g, Φ_i, Γ, A_{22}, C_2, U_d, and U_b to capture the preceding version of Rosen, Murphy, and Scheinkman's (1994) model. Hansen and Sargent (1994) describe the correspondence between partial equilibrium models like Rosen, Murphy, and Scheinkman's and Hansen and Sargent's general equilibrium framework. This involves specifying preferences so that household's first-order conditions can be interpreted as a partial equilibrium demand curve.

Preferences

Set $\Lambda = 0$, $\Delta_h = 0$, $\Theta_h = 0$, $\Pi = \alpha_1^{-1}$, and $b_t = \Pi d_t + \Pi\alpha_0$. With these settings, Hansen and Sargent's marginal condition for the household's problem becomes

$$c_t = \Pi^{-1} b_t - \Pi^{-1} p_t,$$

or

$$c_t = \alpha_0 - \alpha_1 p_t + d_t,$$

thus delivering the appropriate demand curve (i.e., Eq. (116)).

Technology

The law of motion for capital is

$$\begin{bmatrix} x_t \\ x_{t-1} \\ x_{t-2} \end{bmatrix} = \begin{bmatrix} (1-\delta) & 0 & g \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ x_{t-3} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} i_t,$$

or

$$k_t = \Delta_k k_{t-1} + \Theta_h i_t.$$

[11] The costs in Eq. (115) are all absent in the work of Rosen, Murphy, and Scheinkman (1994), though such costs are implicitly taken into account by them in motivating their decision to "solve stable roots backwards and unstable roots forwards." To capture Rosen, Murphy, and Scheinkman's solution, we can set each of the ψ_j's to a positive but very small number.

Here $i_t = -c_t$.

We use adjustment costs to capture the holding and slaughtering costs. We set

$$g_{1t} = f_1 x_t + f_2 h_t,$$

or

$$g_{1t} = f_1[(1-\delta)x_{t-1} + g x_{t-3} - c_t] + f_2 h_t.$$

We set

$$\begin{aligned} g_{2t} &= f_3 x_{t-1} + f_4 h_t \\ g_{3t} &= f_5 x_{t-1} + f_6 h_t. \end{aligned}$$

Notice that

$$\begin{aligned} g_{1t}^2 &= f_1^2 x_t^2 + f_2 h_t^2 + 2 f_1 f_2 x_t h_t \\ g_{2t}^2 &= f_3^2 x_{t-1}^2 + f_2 h_t^2 + 2 f_3 f_4 x_{t-1} h_t \\ g_{3t}^2 &= f_5^2 x_{t-2}^2 + f_6 h_t^2 + 2 f_5 f_6 x_{t-2} t h_t. \end{aligned}$$

Thus, we set

$$f_1^2 = \frac{\psi_1}{2} \quad f_2^2 = \frac{\psi_2}{2} \quad f_3^2 = \frac{\psi_3}{2}$$
$$2 f_1 f_2 = 1 \quad 2 f_3 f_4 = \gamma_1 g \quad 2 f_5 f_6 = \gamma_0 g.$$

To capture the feeding costs we set

$$g_{4t} = f_7 c_t + f_8 m_t$$

and set

$$f_7^2 = \frac{\psi_4}{2} \quad 2 f_7 f_8 = 1.$$

Thus, we set

$$\begin{bmatrix} 1 \\ f_1 \\ 0 \\ 0 \\ -f_7 \end{bmatrix} c_t + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} i_t + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} g_{1t} \\ g_{2t} \\ g_{3t} \\ g_{4t} \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 & 0 \\ f_1(1-\delta) & 0 & g f_1 \\ f_3 & 0 & 0 \\ 0 & f_5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ x_{t-3} \end{bmatrix} + \begin{bmatrix} 0 \\ f_2 h_t \\ f_4 h_t f_6 h_t \\ m_t \end{bmatrix}.$$

We also set $d_t = U_d z_t$, where

$$U_d = \begin{bmatrix} 0 \\ f_2 U_h \\ f_4 U_h \\ f_6 U_h \\ f_8 U_m \end{bmatrix},$$

Table 1. Parameter estimates for "Cattle Cycle" example.

Parameters	Estimates	Standard Errors
α_0	146	33.4
α_1	1.27	0.323
γ_0	0.647	11.5
γ_1	1.77	12.0
g	0.938	0.0222
ρ_h	0.888	0.115
ρ_m	0.699	0.0417
σ_h	6.82	10.6
σ_m	4.04	1.05
σ_y	0.273	0.0383
σ_c	4.82	0.531

where $[U_h, U_m]$ are selector vectors that pick off h_t and m_t from the exogenous state vector z_t. The vector z_t is assumed to be a vector autoregressive process. We assume that the processes for d_t, h_t, and m_t are given by

$$\begin{aligned} d_{t+1} &= \rho_d d_t + \epsilon_{d,t}, \\ h_{t+1} &= (1-\rho_h)\mu_h + \rho_h h_t + \epsilon_{h,t}, \\ m_{t+1} &= (1-\rho_m)\mu_m + \rho_m m_t + \epsilon_{m,t}, \end{aligned}$$

where $E\epsilon_{d,t}^2 = \sigma_d^2$, $E\epsilon_{h,t}^2 = \sigma_h^2$, and $E\epsilon_{m,t}^2 = \sigma_m^2$. The disturbances $\epsilon_{d,t}$, $\epsilon_{h,t}$, and $\epsilon_{m,t}$ are white noise processes that are uncorrelated at all lags.

To compute parameter estimates, we use the same data set as Rosen, Murphy, and Scheinkman (1994) which includes annual observations for y_t, c_t, and p_t for the United States during the period 1900-1990.[12] We assume that there is error in measuring the total stock of cattle, y_t, and the slaughter rate, c_t. In particular, we assume that the (1,1) element of R, the variance-covariance matrix of the measurement errors, is equal to σ_y^2, and we assume that the (2,2) element of R is equal to σ_c^2. All other elements of R are set equal to zero.

We are now equipped to estimate the parameters of this economy by applying the formulas of the previous sections. We start with some a priori restrictions. Assume that $\beta = 0.96$, $\delta = 0$, $f_j = 0.0001$, $j = 1,3,5,7$, $\rho_d = 0$, $\sigma_d = 0$, $\mu_h = 37$, and $\mu_m = 63$. The remaining parameters are elements of Θ, i.e., $\Theta = [\alpha_0,\ \alpha_1,\ \gamma_0,\ \gamma_1,\ \rho_h,\ \rho_m,\ \sigma_h,\ \sigma_m,\ \sigma_y,\ \sigma_c]$. In Table 1, we report estimates of these parameters and standard errors for the estimates. Note that from the values for α_0 and α_1 we can get an estimate of the demand elasticity. For this model, the elasticity is given by -0.61.[13] The val-

[12] The sources of this data are United States, Bureau of the Census (1975) and (1989). y is the total stock of cattle excluding milk cows, c is the cattle slaughtered, and p is the price of slaughtered cattle.

[13] This estimate is $\alpha_1 \times p_0/c_0$ (-1.27×0.48).

ues of γ_0 and γ_1 give us information about the holding costs. The estimates indicate that the costs are higher for calves than for yearlings. However, the standard errors on γ_0 and γ_1 indicate that these parameters are not precisely estimated. The value of g implies that $0.94 x_{t-1}$ calves are born at date t, where x_{t-1} is the breeding stock at $t-1$. This estimate is higher than Rosen, Murphy, and Scheinkman's (1994) estimate of 0.85. The estimates of ρ_h and ρ_m imply that there is persistence in the processes for holding and feeding costs. Finally, the estimates of σ_y and σ_c indicate that the measurement error is higher for the slaughter rate than for the total stock.

In Figures 1 through 3, we plot the predicted and actual time series for the stock of cattle, the slaughter rate, and the price. The predicted series are the one-step-ahead forecasts, e.g., $\bar{G}\hat{x}_t$. These plots support the claim of Rosen, Murphy, and Scheinkman (1994) that the model does well in capturing the cyclical fluctuations in the cattle market.

Conclusion

We have consigned perhaps the most useful parts of this paper to the appendixes, which contain formulas for computing $\frac{\partial L_t}{\partial \theta}$. Resort to these formulas can be avoided by using numerical derivatives, as was done for example by İmrohoroğlu (1993). However, for problems with sizable numbers of parameters, these formulas are very valuable. In terms of consequence for speed of the computations, the decision whether or not to use these formulas as against numerical derivatives will dwarf the choice of a particular equilibrium computation algorithm.

Appendix A: Computing $\frac{\partial L}{\partial \theta}$ and $\frac{\partial L_t}{\partial \theta}$ for a state-space model

Differentiating the log-likelihood function with respect to the free parameters of the economic model can be broken into two steps: first, differentiating the log-likelihood function with respect to matrices appearing in the state-space model (102); and second, differentiating the parameters of the state-space model (98) with respect to the free parameters of the underlying economic model. In this appendix, we derive $\frac{\partial L}{\partial \theta}$ in terms of the derivatives of A_o, C, G, D, R, $\hat{x}_0$, Σ_0, and $\{z_t, t=0,\dots T\}$. We ignore the Jacobian in Eq. (108) since it differs for each problem. In Appendix B, we show how to compute derivatives of A_o and C for the linear-quadratic and nonlinear economies with and without distortions.

The formula for $\frac{\partial L}{\partial \theta}$

For the first step, we take as given A_o, C, G, D, R, $\hat{x}_0$, Σ_0, and $\{z_t, t=0,\dots T\}$, and their derivatives with respect to the deeper economic parameters. We shall show that the derivative of the log-likelihood function is

$$
\begin{aligned}
\frac{\partial L}{\partial\theta} = \sum_{t=0}^{T-1} \big[& 2 \text{ trace}\{\frac{\partial A_o}{\partial\theta}\Sigma_t\bar{G}'M_tG - \hat{x}_tu_t'\Omega_t^{-1}G\} + 2 \text{ trace}\{\frac{\partial C}{\partial\theta}C'G'M_tG\} \\
& + 2 \text{ trace}\{\frac{\partial G}{\partial\theta}(Ao\Sigma_t\bar{G}'M_t - \Sigma_t\bar{G}'M_tD + CC'G'M_t - A_o\hat{x}_tu_t'\Omega_t^{-1} \\
& \qquad + \hat{x}_tu_t'\Omega_t^{-1}D)\} \\
& - 2 \text{ trace}\{\frac{\partial D}{\partial\theta}G\Sigma_t\bar{G}'M_t - z_tu_t'\Omega_t^{-1} + G\hat{x}_tu_t'\Omega_t^{-1})\} \\
& + \text{ trace}\{\frac{\partial R}{\partial\theta}M_t\} + \text{ trace}\{\frac{\partial\Sigma_t}{\partial\theta}\bar{G}'M_t\bar{G}\} - 2 \text{ trace}\{\frac{\partial\hat{x}_t}{\partial\theta}u_t'\Omega_t^{-1}\bar{G}\} \\
& + 2 \text{ trace}\{\frac{\partial z_{t+1}}{\partial\theta}u_t'\Omega_t^{-1}\} - 2 \text{ trace}\{\frac{\partial z_t}{\partial\theta}u_t'\Omega_t^{-1}D\}\big], \qquad (117)
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial\Sigma_{t+1}}{\partial\theta} = & \frac{\partial A_o}{\partial\theta}\Sigma_tA_o' + A_o\frac{\partial\Sigma_t}{\partial\theta}A_o' + A_o\Sigma_t\frac{\partial A_o}{\partial\theta}' + \frac{\partial C}{\partial\theta}C' + C\frac{\partial C'}{\partial\theta} \\
& - \Big(\frac{\partial C}{\partial\theta}C'G' + C\frac{\partial C'}{\partial\theta}G' + CC'\frac{\partial G'}{\partial\theta} + \frac{\partial A_o}{\partial\theta}\Sigma_t\bar{G}' \\
& + A_o\frac{\partial\Sigma_t}{\partial\theta}\bar{G}' + A_o\Sigma_t\frac{\partial\bar{G}'}{\partial\theta}\Big)K_t' + K_t\frac{\partial\Omega_t}{\partial\theta}K_t' \\
& - K\Big(\frac{\partial\bar{G}}{\partial\theta}\Sigma_tA_o' + \bar{G}\frac{\partial\Sigma_t}{\partial\theta}A_o' + \bar{G}\Sigma_t\frac{\partial A_o}{\partial\theta}' \\
& + \frac{\partial G}{\partial\theta}CC' + G\frac{\partial C}{\partial\theta}C' + GC\frac{\partial C'}{\partial\theta}\Big) \qquad (118) \\
\frac{\partial\hat{x}_{t+1}}{\partial\theta} = & \bar{A}_o\frac{\partial\hat{x}_t}{\partial\theta} + (\frac{\partial A_o}{\partial\theta} - \frac{\partial K_t}{\partial\theta}\bar{G} - K_t\frac{\partial\bar{G}}{\partial\theta})\hat{x}_t + \frac{\partial K_t}{\partial\theta}\bar{z}_t \\
& + K_t(\frac{\partial z_{t+1}}{\partial\theta} - D\frac{\partial z_t}{\partial\theta}). \qquad (119)
\end{aligned}
$$

The expressions in (118) and (119) follow from the definitions of Σ_t in Eq. (101) and $\hat{x}_t$ in Eq. (102). The initial conditions, $\hat{x}_0$ and Σ_0, and their derivatives are assumed to be given.

If Σ_0 is given by the steady-state solution of the Riccati equation, then the computation can be simplified. The formula for the derivative of the log-likelihood function is given by

$$
\begin{aligned}
\frac{\partial L}{\partial\theta} = 2T \text{ trace}\{\frac{\partial A_o}{\partial\theta} & (\Sigma\bar{G}'MG - \Gamma_{\hat{x}u}\Omega^{-1}G - \Gamma_{\hat{x}\lambda}(I - KG) \\
& - \Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}(I - KG) - \Sigma\bar{A}_o'\Gamma_{u\lambda}'\Omega^{-1}G \\
& + \Sigma\bar{A}_o'\Pi(I - KG))\}
\end{aligned}
$$

$$
\begin{aligned}
&+ 2T \text{ trace}\{\frac{\partial C}{\partial \theta} C'\big(G'MG - G'\Omega^{-1}\Gamma_{u\lambda}(I-KG) \\
&\qquad - (I-G'K')\Gamma'_{u\lambda}\Omega^{-1}G + (I-G'K')\Pi(I-KG)\big)\} \\
&+ 2T \text{ trace}\{\frac{\partial G}{\partial \theta}(A_o\Sigma\bar{G}'M - \Sigma\bar{G}'MD + CC'G'M \\
&\qquad - A_o\Gamma_{\hat{x}u}\Omega^{-1} + \Gamma_{\hat{x}u}\Omega^{-1}D + A_o\Gamma_{\hat{x}\lambda}K \\
&\qquad - \Gamma_{\hat{x}\lambda}KD - CC'(I-G'K')\Gamma'_{u\lambda}\Omega^{-1} \\
&\qquad + CC'G'\Omega^{-1}\Gamma_{u\lambda}K - A_o\Sigma\bar{A}'_o\Gamma'_{u\lambda}\Omega^{-1} \\
&\qquad + \Sigma\bar{A}'_o\Gamma'_{u\lambda}\Omega^{-1}D + A_o\Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}K \\
&\qquad - \Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}KD - A_o\Sigma\bar{A}'_o\Pi K + \Sigma\bar{A}'_o\Pi KD \\
&\qquad - CC'\Pi K + CC'G'K'\Pi K)\} \\
&- 2T \text{ trace}\{\frac{\partial D}{\partial \theta}(G\Sigma\bar{G}'M + (\Gamma_{zu} - G\Gamma_{\hat{x}u})\Omega^{-1} \\
&\qquad + G\Gamma_{\hat{x}\lambda}K - \Gamma_{z\lambda}K - G\Sigma\bar{A}'_o\Gamma'_{u\lambda}\Omega^{-1} \\
&\qquad + G\Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}K - G\Sigma\bar{A}'_o\Pi K)\} \\
&+ 2T \text{ trace}\{\frac{\partial R}{\partial \theta}(\frac{1}{2}M + \Omega^{-1}\Gamma_{u\lambda}K + \frac{1}{2}K'\Pi K)\} \\
&+ 2 \text{ trace}\{\sum_{t=0}^{T-1}(\frac{\partial z_{t+1}}{\partial \theta} - D\frac{\partial z_t}{\partial \theta})u'_t\Omega^{-1}\} \\
&- 2 \text{ trace}\{\sum_{t=1}^{T-1}(\frac{\partial z_t}{\partial \theta} - D\frac{\partial z_{t-1}}{\partial \theta})\lambda'_t K\} - 2 \text{ trace}\{\frac{\partial \hat{x}_0}{\partial \theta}\lambda'_0\}, \qquad (120)
\end{aligned}
$$

where Σ is the asymptotic state covariance matrix found by iterating on Eq. (101) and $\bar{G}$, K, Ω, u_t and $\hat{x}_t$ are defined in Eqs. (98), (99), (100), and (102), and

$$
\begin{aligned}
\lambda_t &= (A_o - K\bar{G})'\lambda_{t+1} + \bar{G}'\Omega^{-1}u_t, \qquad t = 0, \ldots, T-2 \\
\lambda_{T-1} &= \bar{G}'\Omega^{-1}u_{T-1} \\
\Gamma_{uu} &= \frac{1}{T}\sum_{t=0}^{T-1} u_t u'_t \\
\Gamma_{\hat{x}u} &= \frac{1}{T}\sum_{t=0}^{T-1} \hat{x}_t u'_t \\
\Gamma_{zu} &= \frac{1}{T}\sum_{t=0}^{T-1} z_t u'_t
\end{aligned}
$$

$$\Gamma_{u\lambda} = \frac{1}{T}\sum_{t=1}^{T-1} u_{t-1}\lambda_t' \tag{121}$$

$$\Gamma_{\hat{x}\lambda} = \frac{1}{T}\sum_{t=1}^{T-1} \hat{x}_{t-1}\lambda_t' \tag{122}$$

$$\Gamma_{z\lambda} = \frac{1}{T}\sum_{t=1}^{T-1} z_{t-1}\lambda_t' \tag{123}$$

$$\begin{aligned} M &= \Omega^{-1} - \Omega^{-1}\Gamma_{uu}\Omega^{-1} \\ \bar{A}_o &= A_o - K\bar{G} \\ \Pi &= \bar{A}_o'\Pi\bar{A}_o + \bar{G}'M\bar{G} - \bar{G}'\Omega^{-1}\Gamma_{u\lambda}\bar{A}_o - \bar{A}_o'\Gamma_{u\lambda}'\Omega^{-1}\bar{G}. \end{aligned}$$

In the remainder of this appendix, we derive the formulas in Eq. (117) and Eq. (120). Readers who are not interested in this derivation can skip the rest of this appendix.

Derivation of the formulas

The derivative of the log-likelihood function with respect to any element θ of the parameter vector is given by

$$\begin{aligned} \frac{\partial L}{\partial \theta} &= \sum_{t=0}^{T-1} \mathrm{trace}\{\frac{\partial \Omega_t}{\partial \theta}M_t\} \; + \sum_{t=0}^{T-1} \mathrm{trace}\{(\frac{\partial u_t}{\partial \theta}u_t' + u_t\frac{\partial u_t'}{\partial \theta})\Omega_t^{-1}\} \\ &= S_1 \qquad\qquad\qquad\qquad + S_2, \end{aligned} \tag{124}$$

where $M_t = \Omega_t^{-1} - \Omega_t^{-1}u_t u_t'\Omega_t^{-1}$ and $\Omega_t = Eu_t u_t'$. We start with the first term in the expression for the derivative of the log-likelihood function, S_1. For this, we need the derivative of the covariance matrix, Ω_t, which satisfies

$$\begin{aligned} \frac{\partial \Omega_t}{\partial \theta} &= \frac{\partial \bar{G}}{\partial \theta}\Sigma_t\bar{G}' + \bar{G}\frac{\partial \Sigma_t}{\partial \theta}\bar{G}' + \bar{G}\Sigma_t\frac{\partial \bar{G}}{\partial \theta}' + \frac{\partial R}{\partial \theta} + \frac{\partial G}{\partial \theta}CC'G' \\ &\quad + G\frac{\partial C}{\partial \theta}C'G' + GC\frac{\partial C}{\partial \theta}'G' + GCC'\frac{\partial G}{\partial \theta}' \\ &= (\frac{\partial G}{\partial \theta}A_o + G\frac{\partial A_o}{\partial \theta} - \frac{\partial D}{\partial \theta}G - D\frac{\partial G}{\partial \theta})\Sigma_t\bar{G}' + \bar{G}\frac{\partial \Sigma_t}{\partial \theta}\bar{G}' \\ &\quad + \bar{G}\Sigma_t(A_o'\frac{\partial G}{\partial \theta}' + \frac{\partial A_o}{\partial \theta}'G' - G'\frac{\partial D}{\partial \theta}' - \frac{\partial G}{\partial \theta}'D') + \frac{\partial R}{\partial \theta} \\ &\quad + \frac{\partial G}{\partial \theta}CC'G' + G\frac{\partial C}{\partial \theta}C'G' + GC\frac{\partial C}{\partial \theta}'G' + GCC'\frac{\partial G}{\partial \theta}'. \end{aligned} \tag{125}$$

The second equality follows from the definition of $\bar{G}$. If we post-multiply the derivative of Ω_t by M_t and take the trace of the result, we have the first term of the derivative of the log-likelihood function in Eq. (124):

$$S_1 = \sum_{t=0}^{T-1}[2\ \mathrm{trace}(\frac{\partial A_o}{\partial \theta}\Sigma_t\bar{G}'M_tG) + 2\ \mathrm{trace}(\frac{\partial C}{\partial \theta}C'G'M_tG)$$

$$+ 2\ \mathrm{trace}(\frac{\partial G}{\partial \theta}\{A_o\Sigma_t\bar{G}'M_t - \Sigma_t\bar{G}'M_tD + CC'G'M_t\})$$
$$- 2\ \mathrm{trace}(\frac{\partial D}{\partial \theta}G\Sigma_t\bar{G}'M_t) + \mathrm{trace}(\frac{\partial R}{\partial \theta}M_t)$$
$$+ \mathrm{trace}(\frac{\partial \Sigma_t}{\partial \theta}\bar{G}'M_t\bar{G})\big]. \tag{126}$$

Note that the formula for S_1 depends on derivatives $\frac{\partial A_o}{\partial \theta}$, $\frac{\partial C}{\partial \theta}$, $\frac{\partial G}{\partial \theta}$, $\frac{\partial D}{\partial \theta}$, and $\frac{\partial R}{\partial \theta}$, which are known, and $\frac{\partial \Sigma_t}{\partial \theta}$, which is yet to be derived.

We now turn to the second term of the log-likelihood function derivative, $S_2 = \mathrm{trace}(\partial u_t u_t'/\partial\theta\Omega_t^{-1})$. Let $\Gamma_{uu}(t) = u_tu_t'$. By definition, $\Gamma_{uu}(t) = (\bar{z}_t - \bar{G}\hat{x}_t)(\bar{z}_t - \bar{G}\hat{x}_t)'$; therefore, its derivative is given by

$$\begin{aligned}
\frac{\partial \Gamma_{uu}(t)}{\partial \theta} &= (\frac{\partial \bar{z}_t}{\partial \theta} - \frac{\partial \bar{G}}{\partial \theta}\hat{x}_t - \bar{G}\frac{\partial \hat{x}_t}{\partial \theta})u_t' + u_t(\frac{\partial \bar{z}_t}{\partial \theta} - \frac{\partial \bar{G}}{\partial \theta}\hat{x}_t - \bar{G}\frac{\partial \hat{x}_t}{\partial \theta})' \\
&= (\frac{\partial z_{t+1}}{\partial \theta} - \frac{\partial D}{\partial \theta}z_t - D\frac{\partial z_t}{\partial \theta} - \frac{\partial G}{\partial \theta}A_o\hat{x}_t - G\frac{\partial A_o}{\partial \theta}\hat{x}_t \\
&+ \frac{\partial D}{\partial \theta}G\hat{x}_t + D\frac{\partial G}{\partial \theta}\hat{x}_t - \bar{G}\frac{\partial \hat{x}_t}{\partial \theta})u_t' \\
&+ u_t(\frac{\partial z_{t+1}}{\partial \theta}' - z_t'\frac{\partial D'}{\partial \theta} - \frac{\partial z_t}{\partial \theta}'D' - \hat{x}_t'A_o'\frac{\partial G'}{\partial \theta} - \hat{x}_t'\frac{\partial A_o}{\partial \theta}'G' \\
&+ \hat{x}_t'G'\frac{\partial D'}{\partial \theta} + \hat{x}_t'\frac{\partial G'}{\partial \theta}D' - \frac{\partial \hat{x}_t}{\partial \theta}'\bar{G}').
\end{aligned} \tag{127}$$

If we post-multiply this derivative by Ω_t^{-1}, take the trace of the resulting matrix, and sum over t, then we have the second term of the derivative of the log-likelihood function, i.e.,

$$\begin{aligned}
S_2 = -\sum_{t=0}^{T-1}\big[&2\ \mathrm{trace}\{\frac{\partial A_o}{\partial \theta}\hat{x}_tu_t'\Omega_t^{-1}G\} + 2\ \mathrm{trace}\{\frac{\partial G}{\partial \theta}(A_o\hat{x}_tu_t'\Omega_t^{-1} - \hat{x}_tu_t'\Omega_t^{-1}D)\} \\
&+ 2\ \mathrm{trace}\{\frac{\partial D}{\partial \theta}(z_tu_t' - G\hat{x}_tu_t')\Omega_t^{-1}\} - 2\ \mathrm{trace}\{\sum_{t=0}^{T-1}\frac{\partial z_{t+1}}{\partial \theta}u_t'\Omega_t^{-1}\} \\
&+ 2\ \mathrm{trace}\{\sum_{t=0}^{T-1}\frac{\partial z_t}{\partial \theta}u_t'\Omega_t^{-1}D\} + 2\ \mathrm{trace}\{\sum_{t=0}^{T-1}\frac{\partial \hat{x}_t}{\partial \theta}u_t'\Omega_t^{-1}\bar{G}\}\big].
\end{aligned} \tag{128}$$

Sum the expressions in Eqs. (126) and (128) to get the expression for the derivative of the log-likelihood function in (117).

For the time-invariant case, several more steps are needed. First, we derive the last term in Eq. (128) in terms of the derivatives that are taken as inputs. To simplify notation, we first define the sequences $\{d_t\}$ and $\{\lambda_t\}$ as follows:

$$\begin{aligned}
d_t &= (\frac{\partial A_o}{\partial \theta} - \frac{\partial K}{\partial \theta}\bar{G} - K\frac{\partial \bar{G}}{\partial \theta})\hat{x}_t + \frac{\partial K}{\partial \theta}\bar{z}_t + K\frac{\partial \bar{z}_t}{\partial \theta}, \quad && t = 0, \dots, T-1 \\
\lambda_t &= (A_o - K\bar{G})'\lambda_{t+1} + \bar{G}'\Omega^{-1}u_t, && t = 0, \dots, T-2 \\
\lambda_{T-1} &= \bar{G}'\Omega^{-1}u_{T-1}.
\end{aligned} \tag{129}$$

Notice that the time subscripts have been dropped from K and Ω since the time-invariant case assumes that $\Sigma_t = \Sigma$ for all t. Let $\bar{A}_o = A_o - K\bar{G}$. Notice that since $\hat{x}_{t+1} = \bar{A}_o\hat{x}_t + K\bar{z}_t$, its derivative is given by

$$\frac{\partial \hat{x}_{t+1}}{\partial \theta} = \bar{A}_o \frac{\partial \hat{x}_t}{\partial \theta} + d_t. \tag{130}$$

Write out the last term in Eq. (128) and substitute in $\hat{x}_t = \bar{A}_o^t + \sum_{s=0}^{t-1} \bar{A}_o^{s-1} \times d_{t-s}$. Then group terms involving $\hat{x}_0$ and d_t, $t = 0, \ldots T-2$. These steps lead to

$$\begin{aligned}
-\frac{2}{T}\text{trace}(\sum_{t=0}^{T-1} \frac{\partial \hat{x}_t}{\partial \theta} u_t' \Omega^{-1}\bar{G}) &= -\frac{2}{T}\text{trace}(\frac{\partial \hat{x}_0}{\partial \theta}\lambda_0' + \sum_{t=1}^{T-1} d_{t-1}\lambda_t') \\
&= -\frac{2}{T}\ \text{trace}(\frac{\partial \hat{x}_0}{\partial \theta}\lambda_0') - 2\ \text{trace}\{(\frac{\partial A_o}{\partial \theta} - \frac{\partial K}{\partial \theta}\bar{G} - K\frac{\partial G}{\partial \theta}A_o \\
&\quad - KG\frac{\partial A_o}{\partial \theta} + K\frac{\partial D}{\partial \theta}G + KD\frac{\partial G}{\partial \theta})\Gamma_{\hat{x}\lambda} + \frac{\partial K}{\partial \theta}\Gamma_{\bar{z}\lambda} \\
&\quad + \frac{1}{T}K\sum_{t=1}^{T-1} \frac{\partial z_t}{\partial \theta}\lambda_t' - K\frac{\partial D}{\partial \theta}\Gamma_{z\lambda} - \frac{1}{T}KD\sum_{t=1}^{T-1}\frac{\partial z_{t-1}}{\partial \theta}\lambda_t'\} \\
&= -2\ \text{trace}\{\frac{\partial A_o}{\partial \theta}\Gamma_{\hat{x}\lambda}(I - KG)\} \\
&\quad + 2\ \text{trace}\{\frac{\partial G}{\partial \theta}(A_o\Gamma_{\hat{x}\lambda}K - \Gamma_{\hat{x}\lambda}KD\} \\
&\quad - 2\ \text{trace}\{\frac{\partial D}{\partial \theta}(G\Gamma_{\hat{x}\lambda}K - \Gamma_{z\lambda}K)\} \\
&\quad - \frac{2}{T}\ \text{trace}\{K\sum_{t=1}^{T-1}\frac{\partial z_t}{\partial \theta}\lambda_t' - KD\sum_{t=1}^{T-1}\frac{\partial z_{t-1}}{\partial \theta}\lambda_t'\} \\
&\quad - \frac{2}{T}\ \text{trace}(\frac{\partial \hat{x}_0}{\partial \theta}\lambda_0') - 2\ \text{trace}\{\frac{\partial K}{\partial \theta}\Gamma_{u\lambda}\},
\end{aligned} \tag{131}$$

where $\Gamma_{u\lambda}$, $\Gamma_{\hat{x}\lambda}$, and $\Gamma_{z\lambda}$ are the sums defined in Eqs. (121) through Eq. (123) and $\Gamma_{\bar{z}\lambda} = \sum_{t=1}^{T-1} \bar{z}_{t-1}\lambda_t'/T$. The second equality follows from the definitions of d_{t-1} and $\bar{G}$ and some algebraic manipulation. The last term in Eq. (131) uses the fact that $u_t = \bar{z}_t - \bar{G}\hat{x}_t$. With the exception of $\frac{\partial K}{\partial \theta}$, the expression in Eq. (131) is a function of known derivatives. The expression for $\frac{\partial K}{\partial \theta}$ follows from the definition in Eq. (99) and is given by

$$\begin{aligned}
\frac{\partial K}{\partial \theta} = \Bigg[&\frac{\partial C}{\partial \theta}C'G' + C\frac{\partial C'}{\partial \theta}G' + CC'\frac{\partial G'}{\partial \theta} + \frac{\partial A_o}{\partial \theta}\Sigma\bar{G}' + A_o\frac{\partial \Sigma}{\partial \theta}\bar{G}' \\
&+ A_o\Sigma A_o'\frac{\partial G'}{\partial \theta} + A_o\Sigma\frac{\partial A_o{}'}{\partial \theta}G' - A_o\Sigma G'\frac{\partial D'}{\partial \theta} - A_o\Sigma\frac{\partial G'}{\partial \theta}D'\Bigg]\Omega^{-1}
\end{aligned}$$

$$- (CC'G' + A_o\Sigma\bar{G}')\Omega^{-1}\Big[\frac{\partial G}{\partial\theta}A_o\Sigma\bar{G}' + G\frac{\partial A_o}{\partial\theta}\Sigma\bar{G}' - \frac{\partial D}{\partial\theta}G\Sigma\bar{G}'$$
$$- D\frac{\partial G}{\partial\theta}\Sigma\bar{G}' + \bar{G}\frac{\partial\Sigma}{\partial\theta}\bar{G}' + \bar{G}\Sigma A_o'\frac{\partial G'}{\partial\theta} + \bar{G}\Sigma\frac{\partial A_o}{\partial\theta}'G'$$
$$- \bar{G}\Sigma G'\frac{\partial D'}{\partial\theta} - \bar{G}\Sigma\frac{\partial G'}{\partial\theta}D' + \frac{\partial R}{\partial\theta} + \frac{\partial G}{\partial\theta}CC'G'$$
$$+ G\frac{\partial C}{\partial\theta}C'G' + GC\frac{\partial C'}{\partial\theta}G' + GCC'\frac{\partial G'}{\partial\theta}\Big]\Omega^{-1}. \tag{132}$$

Note that we have written $\frac{\partial\bar{G}}{\partial\theta}$ in terms of $\frac{\partial G}{\partial\theta}$, $\frac{\partial A_o}{\partial\theta}$, and $\frac{\partial D}{\partial\theta}$. Substituting $\frac{\partial K}{\partial\theta}$ into the expression in Eq. (131) and rearranging terms, we have

$$-\frac{2}{T}\text{trace}(\sum_{t=0}^{T-1}\frac{\partial\hat{x}_t}{\partial\theta}u_t'\Omega^{-1}\bar{G}) =$$
$$- 2\ \text{trace}\{\frac{\partial A_o}{\partial\theta}(\Gamma_{\hat{x}\lambda}(I-KG) + \Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}(I-KG)$$
$$+ \Sigma\bar{A}_o'\Gamma_{u\lambda}'\Omega^{-1}G\}$$
$$- 2\ \text{trace}\{\frac{\partial C}{\partial\theta}C'(G'\Omega^{-1}\Gamma_{u\lambda}(I-KG) + (I-G'K')\Gamma_{u\lambda}'\Omega^{-1}G)\}$$
$$+ 2\ \text{trace}\{\frac{\partial G}{\partial\theta}(A_o\Gamma_{\hat{x}\lambda}K - \Gamma_{\hat{x}\lambda}KD - CC'(I-G'K')\Gamma_{u\lambda}'\Omega^{-1}$$
$$+ CC'G'\Omega^{-1}\Gamma_{u\lambda}K - A_o\Sigma\bar{A}_o'\Gamma_{u\lambda}'\Omega^{-1} + \Sigma\bar{A}_o'\Gamma_{u\lambda}'\Omega^{-1}D$$
$$+ A_o\Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}K - \Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}KD)\}$$
$$- 2\ \text{trace}\{\frac{\partial D}{\partial\theta}(G\Gamma_{\hat{x}\lambda}K - \Gamma_{z\lambda}K - G\Sigma\bar{A}_o'\Gamma_{u\lambda}'\Omega^{-1} + G\Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}K)\}$$
$$+ 2\ \text{trace}\{\frac{\partial R}{\partial\theta}\Omega^{-1}\Gamma_{u\lambda}K\}$$
$$- \frac{2}{T}\ \text{trace}\{K\sum_{t=1}^{T-1}\frac{\partial z_t}{\partial\theta}\lambda_t' - KD\sum_{t=1}^{T-1}\frac{\partial z_{t-1}}{\partial\theta}\lambda_t'\}$$
$$- \frac{2}{T}\ \text{trace}\{\frac{\partial\hat{x}_0}{\partial\theta}\lambda_0'\} - 2\ \text{trace}\{\frac{\partial\Sigma}{\partial\theta}(\bar{G}'\Omega^{-1}\Gamma_{u\lambda}\bar{A}_o)\}. \tag{133}$$

Therefore, the expression for the second term of the log-likelihood function derivative, S_2, is given by

$$S_2 = - 2\ \text{trace}\{\frac{\partial A_o}{\partial\theta}(\Gamma_{\hat{x}u}\Omega^{-1}G + \Gamma_{\hat{x}\lambda}(I-KG) + \Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}(I-KG)$$
$$+ \Sigma\bar{A}_o'\Gamma_{u\lambda}'\Omega^{-1}G\}$$

$$
\begin{aligned}
&- 2\ \mathrm{trace}\{\frac{\partial C}{\partial \theta} C'(G'\Omega^{-1}\Gamma_{u\lambda}(I-KG) + (I-G'K')\Gamma'_{u\lambda}\Omega^{-1}G)\} \\
&- 2\ \mathrm{trace}\{\frac{\partial G}{\partial \theta}(A_o\Gamma_{\hat{x}u}\Omega^{-1} - \Gamma_{\hat{x}u}\Omega^{-1}D - A_o\Gamma_{\hat{x}\lambda}K + \Gamma_{\hat{x}\lambda}KD \\
&\qquad + CC'(I-G'K')\Gamma'_{u\lambda}\Omega^{-1} - CC'G'\Omega^{-1}\Gamma_{u\lambda}K \\
&\qquad + A_o\Sigma\bar{A}'_o\Gamma'_{u\lambda}\Omega^{-1} - \Sigma\bar{A}'_o\Gamma'_{u\lambda}\Omega^{-1}D \\
&\qquad - A_o\Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}K + \Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}KD)\} \\
&- 2\ \mathrm{trace}\{\frac{\partial D}{\partial \theta}((\Gamma_{zu} - G\Gamma_{\hat{x}u})\Omega^{-1} + G\Gamma_{\hat{x}\lambda}K - \Gamma_{z\lambda}K - G\Sigma\bar{A}'_o\Gamma'_{u\lambda}\Omega^{-1} \\
&\qquad + G\Sigma\bar{G}'\Omega^{-1}\Gamma_{u\lambda}K)\} \\
&+ 2\ \mathrm{trace}\{\frac{\partial R}{\partial \theta}\Omega^{-1}\Gamma_{u\lambda}K\} \\
&+ \frac{2}{T}\ \mathrm{trace}\{\sum_{t=0}^{T-1}\frac{\partial z_{t+1}}{\partial \theta}u'_t\Omega^{-1}\} - \frac{2}{T}\ \mathrm{trace}\{\sum_{t=0}^{T-1}\frac{\partial z_t}{\partial \theta}u'_t\Omega^{-1}D\} \\
&- \frac{2}{T}\ \mathrm{trace}\{K\sum_{t=1}^{T-1}\frac{\partial z_t}{\partial \theta}\lambda'_t - KD\sum_{t=1}^{T-1}\frac{\partial z_{t-1}}{\partial \theta}\lambda'_t\} \\
&- \frac{2}{T}\ \mathrm{trace}\{\frac{\partial \hat{x}_0}{\partial \theta}\lambda'_0\} \\
&- 2\ \mathrm{trace}\{\frac{\partial \Sigma}{\partial \theta}\bar{G}'\Omega^{-1}\Gamma_{u\lambda}\bar{A}_o\}. \qquad (134)
\end{aligned}
$$

Our expressions for S_1 in Eq. (126) and S_2 in Eq. (134) depend on $\frac{\partial A_o}{\partial \theta}$, $\frac{\partial C}{\partial \theta}$, $\frac{\partial G}{\partial \theta}$, $\frac{\partial D}{\partial \theta}$, $\frac{\partial R}{\partial \theta}$, which are known, and $\frac{\partial \Sigma}{\partial \theta}$, which we will now derive. Using the expression in Eq. (118) with $\Sigma_{t+1} = \Sigma_t = \Sigma$, we get

$$
\frac{\partial \Sigma}{\partial \theta} = \bar{A}_o\frac{\partial \Sigma}{\partial \theta}\bar{A}'_o + W + W', \qquad (135)
$$

where

$$
\begin{aligned}
W = &\frac{\partial A_o}{\partial \theta}\Sigma A'_o + \frac{\partial C}{\partial \theta}C' - \frac{\partial C}{\partial \theta}C'G'K' - C\frac{\partial C}{\partial \theta}' G'K' - CC'\frac{\partial G}{\partial \theta}' K' \\
&- \frac{\partial A_o}{\partial \theta}\Sigma\bar{G}'K' - A_o\Sigma A'_o\frac{\partial G}{\partial \theta}' K' - A_o\Sigma\frac{\partial A_o}{\partial \theta}' G'K' \\
&+ A_o\Sigma G'\frac{\partial D}{\partial \theta}' K' + A_o\Sigma\frac{\partial G}{\partial \theta}' D'K' + \frac{1}{2}K\frac{\partial R}{\partial \theta}K' \\
&+ K\frac{\partial G}{\partial \theta}A_o\Sigma\bar{G}'K' + KG\frac{\partial A_o}{\partial \theta}\Sigma\bar{G}'K' - K\frac{\partial D}{\partial \theta}G\Sigma\bar{G}'K' \\
&- KD\frac{\partial G}{\partial \theta}\Sigma\bar{G}'K' + K\frac{\partial G}{\partial \theta}CC'G'K' + KG\frac{\partial C}{\partial \theta}C'G'K'. \qquad (136)
\end{aligned}
$$

The terms W and W' in Eq. (135) include all derivatives but $\frac{\partial \Sigma}{\partial \theta}$. To get the expression in Eq. (136), we substituted the expressions for $\frac{\partial \Omega}{\partial \theta}$ and $\frac{\partial G}{\partial \theta}$ into Eq. (135). Let Π be a symmetric matrix that satisfies

$$\Pi = \bar{A}_o' \Pi \bar{A}_o + \frac{1}{2}(H + H'), \tag{137}$$

where

$$H = \bar{G}' M \bar{G} - 2\bar{G}'\Omega^{-1}\Gamma_{u\lambda}\bar{A}_o. \tag{138}$$

Then,

$$\begin{aligned}
\text{trace}(\frac{\partial \Sigma}{\partial \theta} H) &= \text{trace}\{\frac{\partial \Sigma}{\partial \theta}\frac{1}{2}(H + H')\} \\
&= \text{trace}\{\frac{\partial \Sigma}{\partial \theta}(\Pi - \bar{A}_o'\Pi\bar{A}_o)\} \\
&= \text{trace}\{\frac{\partial \Sigma}{\partial \theta}\Pi\} - \text{trace}\{\bar{A}_o\frac{\partial \Sigma}{\partial \theta}\bar{A}_o'\Pi\} \\
&= \text{trace}\{(\frac{\partial \Sigma}{\partial \theta} - \bar{A}_o\frac{\partial \Sigma}{\partial \theta}\bar{A}_o')\Pi\} \\
&= \text{trace}\{(W + W')\Pi\} \\
&= 2\ \text{trace}\{W\Pi\}.
\end{aligned} \tag{139}$$

If we post-multiply W by Π and take 2 times the trace, then we have an expression for trace($\frac{\partial \Sigma}{\partial \theta}H$) in terms of known derivatives, i.e.,

$$\begin{aligned}
\text{trace}(\frac{\partial \Sigma}{\partial \theta} H) = {} & 2\ \text{trace}\{\frac{\partial A_o}{\partial \theta}\Sigma\bar{A}_o'\Pi(I - KG)\} \\
& + 2\ \text{trace}\{\frac{\partial C}{\partial \theta}C'(I - G'K')\Pi(I - KG)\} \\
& - 2\ \text{trace}\{\frac{\partial G}{\partial \theta}(A_o\Sigma\bar{A}_o'\Pi K - \Sigma\bar{A}_o'\Pi K D + CC'(I - G'K')\Pi K\} \\
& + 2\ \text{trace}\{\frac{\partial D}{\partial \theta}G\Sigma\bar{A}_o'\Pi K\} + \text{trace}\{\frac{\partial R}{\partial \theta}K'\Pi K\}.
\end{aligned} \tag{140}$$

Sum S_1, which appears in Eq. (126) with $\Sigma_t = \Sigma$ and $\Omega_t = \Omega$, and S_2 in (134). Substitute in the expression for trace($\frac{\partial \Sigma}{\partial \theta}H$) from Eq. (140). The result is the derivative of the log-likelihood function which is given in Eq. (120).

Standard errors

After we have computed parameter estimates, we want to compute their standard errors as given in Eq. (110). For this, we need to compute the derivative of

$$L_t(\Theta) = \log|\Omega_t| + u_t'\Omega_t^{-1}u_t$$

with respect to any element θ of the parameter vector.[14] This derivative is given by

$$\begin{aligned}
\frac{\partial L_t}{\partial \theta} &= \mathrm{trace}(\Omega_t^{-1}\frac{\partial \Omega_t}{\partial \theta}) + \frac{\partial u_t}{\partial \theta}' \Omega_t^{-1} u_t + u_t' \Omega_t^{-1} \frac{\partial u_t}{\partial \theta} - u_t' \Omega_t^{-1} \frac{\partial \Omega_t}{\partial \theta} \Omega_t^{-1} u_t \\
&= \mathrm{trace}\{(\Omega_t^{-1} - \Omega_t^{-1} u_t u_t' \Omega_t^{-1}) \frac{\partial \Omega_t}{\partial \theta}\} + \mathrm{trace}\{\frac{\partial u_t}{\partial \theta}' \Omega_t^{-1} u_t + u_t' \Omega_t^{-1} \frac{\partial u_t}{\partial \theta}\} \\
&= \mathrm{trace}\{\frac{\partial \Omega_t}{\partial \theta} M_t\} + \mathrm{trace}\{\Omega_t^{-1} \frac{\partial (u_t u_t')}{\partial \theta}\}, \qquad (141)
\end{aligned}$$

where $M_t = \Omega_t^{-1} - \Omega_t^{-1} u_t u_t' \Omega_t^{-1}$. Above, we calculated $\frac{\partial \Omega_t}{\partial \theta}$ and $\frac{\partial (u_t u_t')}{\partial \theta}$. These expressions are given in Eq. (125) and Eq. (127).

Appendix B: Differentiating the state-space model with respect to economic parameters

In this appendix, we describe how to compute derivatives of A_o and C with respect to the free parameters of an economic model. We do this for four economies: a linear-quadratic economy without distortions; a nonlinear economy without distortions; a linear-quadratic economy with distortions; and a nonlinear economy with distortions. Because we use linear approximations for the nonlinear economies, most of the work is in deriving the formulas for the linear-quadratic economies.

A linear-quadratic economy without distortions

The optimization problem is

$$\begin{aligned}
&\max_{\{u_t\}} \sum_{t=0}^{\infty} \beta^t (x_t' Q x_t + u_t' R u_t + 2 x_t' W u_t) \qquad (142) \\
&\text{subject to } x_{t+1} = A x_t + B u_t + C \epsilon_{t+1},
\end{aligned}$$

where each element of ϵ_t is a random variable that is normally distributed with mean 0 and variance equal to 1. We assume that the matrices Q, R, W, A, B, and C depend on a vector of parameters, Θ. Typically, the number of elements in Θ is small relative to the combined number of elements in these matrices. We also assume that the derivatives of the matrices in Eq. (142) with respect to the elements of Θ are known.

The optimal decision function is given by $u_t = -F x_t$, where

$$F = (R + \beta B' P B)^{-1} (\beta B' P A + W') \qquad (143)$$

for P satisfying

$$P = Q + \beta A' P A - (W + \beta A' P B)(R + \beta B' P B)^{-1} (\beta B' P A + W'). \qquad (144)$$

[14] Note that we are again ignoring the Jacobian since the relationship between z and y differs for each problem.

The law of motion for x in equilibrium is

$$x_{t+1} = A_o x_t + C\epsilon_{t+1}, \quad A_o = A - BF. \tag{145}$$

Therefore, the derivative of A_o with respect to an element of Θ is

$$\frac{\partial A_o}{\partial \theta} = \frac{\partial A}{\partial \theta} - \frac{\partial B}{\partial \theta}F - B\frac{\partial F}{\partial \theta}. \tag{146}$$

The derivatives $\frac{\partial A}{\partial \theta}$ and $\frac{\partial B}{\partial \theta}$ depend on the specification of the problem in Eq. (142) and are assumed to be known. The derivative of F is

$$\begin{aligned}\frac{\partial F}{\partial \theta} &= -(R+\beta B'PB)^{-1}\Big(\frac{\partial R}{\partial \theta}+\beta\frac{\partial B'}{\partial \theta}PB+\beta B'\frac{\partial P}{\partial \theta}B+\beta B'P\frac{\partial B}{\partial \theta}\Big)F \\ &\quad + (R+\beta B'PB)^{-1}\Big(\beta\frac{\partial B'}{\partial \theta}PA+\beta B'\frac{\partial P}{\partial \theta}A+\beta B'P\frac{\partial A}{\partial \theta}+\frac{\partial W'}{\partial \theta}\Big).\end{aligned} \tag{147}$$

Notice that this formula depends on the derivative of P, with the remaining derivatives provided by the modeler. The derivative $\frac{\partial P}{\partial \theta}$ satisfies the following equation:

$$\begin{aligned}\frac{\partial P}{\partial \theta} &= \frac{\partial Q}{\partial \theta} + \beta\frac{\partial A'}{\partial \theta}PA + \beta A'\frac{\partial P}{\partial \theta}A + \beta A'P\frac{\partial A}{\partial \theta} - \Big(\frac{\partial W}{\partial \theta} + \beta\frac{\partial A'}{\partial \theta}PB \\ &\quad + \beta A'\frac{\partial P}{\partial \theta}B + \beta A'P\frac{\partial B}{\partial \theta}\Big)F + F'\Big(\frac{\partial R}{\partial \theta} + \beta\frac{\partial B'}{\partial \theta}PB + \beta B'\frac{\partial P}{\partial \theta}B \\ &\quad + \beta B'P\frac{\partial B}{\partial \theta}\Big)F - F'\Big(\beta\frac{\partial B'}{\partial \theta}PA + \beta B'\frac{\partial P}{\partial \theta}A + \beta B'P\frac{\partial A}{\partial \theta} + \frac{\partial W'}{\partial \theta}\Big) \\ &= \beta A_o'\frac{\partial P}{\partial \theta}A_o + \frac{\partial Q}{\partial \theta} + \beta\Big[\frac{\partial A'}{\partial \theta} - F'\frac{\partial B'}{\partial \theta}\Big]PA_o + \beta A_o'P\Big[\frac{\partial A}{\partial \theta} - \frac{\partial B}{\partial \theta}F\Big] \\ &\quad - \frac{\partial W}{\partial \theta}F - F'\frac{\partial W'}{\partial \theta} + F'\frac{\partial R}{\partial \theta}F.\end{aligned} \tag{148}$$

Although this formula determines only an implicit function for $\frac{\partial P}{\partial \theta}$, the gradient of P can be represented explicitly in terms of things we know. Define the gradient operator as follows: for any matrix A that depends on the parameter θ, $\nabla_\theta A = vec(\frac{\partial A}{\partial \theta})$. Then,

$$\begin{aligned}\nabla_\theta P = (I - \beta A_o' \otimes A_o')^{-1}\big\{ &\nabla_\theta Q + \beta(A_o'P \otimes I)\,\nabla_\theta A' + \beta(I \otimes A_o'P)\,\nabla_\theta A \\ &- \beta(A_o'P \otimes F')\,\nabla_\theta B' - \beta(F' \otimes A_o'P)\,\nabla_\theta B - (F' \otimes I)\,\nabla_\theta W \\ &- (I \otimes F')\,\nabla_\theta W' + (F' \otimes F')\,\nabla_\theta R\big\},\end{aligned} \tag{149}$$

which is a function of the gradients of A, B, Q, R, and W. The gradient of P can then be substituted into the following formula for $\nabla_\theta F$

$$\nabla_\theta F = \beta(I \otimes \mathcal{R}B'P)\,\nabla_\theta A - \beta(F' \otimes \mathcal{R}B'P)\,\nabla_\theta B + \beta(A_o'P \otimes \mathcal{R})\,\nabla_\theta B'$$

$$- (F' \otimes \mathcal{R})\,\nabla_\theta R + (I \otimes \mathcal{R})\,\nabla_\theta W' + \beta(A_o' \otimes \mathcal{R}B')\,\nabla_\theta P, \qquad (150)$$

where $\mathcal{R} = (R + \beta B'PB)^{-1}$. Finally, we substitute this expression for $\nabla_\theta F$ into

$$\nabla_\theta A_o = \nabla_\theta A - (F' \otimes I)\,\nabla_\theta B - (I \otimes B)\,\nabla_\theta F. \qquad (151)$$

Since C is chosen by the modeler, we assume that its derivative with respect to θ is known.

A nonlinear economy without distortions

The optimization problem that we start with is

$$\max_{\{u_t\}} E_0 \sum_{t=0}^{\infty} \beta^t r(z_t, \theta) \qquad (152)$$

$$\text{subject to } x_{t+1} = Ax_t + Bu_t + Cw_{t+1}$$

$$z_t = [x_t', u_t']'.$$

We solve a related problem, namely,

$$\max_{\{u_t\}} E_0 \sum_{t=0}^{\infty} \beta^t z_t' M z_t \qquad (153)$$

$$x_{t+1} = Ax_t + Bu_t + Cw_{t+1},$$

where

$$M = e(r(\bar z, \theta) - \frac{\partial r(\bar z, \theta)}{\partial \bar z}' \bar z + \frac{1}{2}\bar z' \frac{\partial^2 r(\bar z, \theta)}{\partial \bar z^2}\bar z)e' + \frac{1}{2}(e\frac{\partial r(\bar z, \theta)}{\partial \bar z}'$$

$$+ \frac{\partial r(\bar z, \theta)}{\partial \bar z}e' - e\bar z' \frac{\partial^2 r(\bar z, \theta)}{\partial \bar z^2} - \frac{\partial^2 r(\bar z, \theta)}{\partial \bar z^2}\bar z e' + \frac{\partial^2 r(\bar z, \theta)}{\partial \bar z^2}), \qquad (154)$$

and where e is a vector of zeros except for a 1 in the element corresponding to the constant term in x_t, $\bar z$ and $\bar w$ are the steady-state values of z_t and w_t, and $S_x = [I_n, 0_{n,k}]$ and $S_u = [0_{k,n}, I_k]$ are selector matrices and imply $z_t = S_x x_t + S_u u_t$, where n is the dimension The latter problem yields the same decision function as that of Eq. (142) (where $Q = S_x' M S_x$, $R = S_u' M S_u$, and $W = S_x' M S_u$).

In the nonlinear case, however, the derivatives are slightly more complicated. To derive $\frac{\partial A_o}{\partial \theta}$, we need to calculate derivatives of the coefficient matrices of the objective function. For this, we need the derivative of M with respect to θ:

$$\begin{aligned}\frac{\partial M}{\partial \theta} = e\Big(&\frac{\partial r(\bar{z},\theta)}{\partial \theta} - \frac{\partial^2 r(\bar{z},\theta)}{\partial \bar{z}\partial\theta}' \bar{z} + \frac{1}{2}\bar{z}'\Big(\nabla_{\bar{z}}\frac{\partial^2 r(\bar{z},\theta)}{\partial \bar{z}^2}\frac{\partial \bar{z}}{\partial \theta}\Big)(:)\,\bar{z} \\ &+ \frac{1}{2}\bar{z}'\frac{\partial^3 r(\bar{z},\theta)}{\partial \bar{z}^2\partial\theta}\bar{z}\Big)e' + \frac{1}{2}\Big(e\frac{\partial^2 r(\bar{z},\theta)}{\partial \bar{z}\partial\theta}' + \frac{\partial^2 r(\bar{z},\theta)}{\partial \bar{z}\partial\theta}e' \\ &- e\bar{z}'\Big(\nabla_{\bar{z}}\frac{\partial^2 r(\bar{z},\theta)}{\partial \bar{z}^2}\frac{\partial \bar{z}}{\partial \theta}\Big)(:) - \Big(\nabla_{\bar{z}}\frac{\partial^2 r(\bar{z},\theta)}{\partial \bar{z}^2}\frac{\partial \bar{z}}{\partial \theta}\Big)(:)\,\bar{z}e' \\ &- e\bar{z}'\frac{\partial^3 r(\bar{z},\theta)}{\partial \bar{z}^2\partial\theta} - \frac{\partial^3 r(\bar{z},\theta)}{\partial \bar{z}^2\partial\theta}\bar{z}e' + \frac{\partial^3 r(\bar{z},\theta)}{\partial \bar{z}^2\partial\theta} \\ &+ \Big(\nabla_{\bar{z}}\frac{\partial^2 r(\bar{z},\theta)}{\partial \bar{z}^2}\frac{\partial \bar{z}}{\partial \theta}\Big)(:)\Big),\end{aligned} \tag{155}$$

where $\nabla_z A(z) = [\frac{\partial}{\partial z_1}A(z), \frac{\partial}{\partial z_2}A(z), \ldots, \frac{\partial}{\partial z_n}A(z)]$ for $A(z)$ which is $n \times n$ and $b(:)$ is an $n \times n$ matrix created from a vector of length n^2 by stacking the first n elements of b into column 1, the next n elements of b into column 2, etc. As this formula indicates, the modeler must provide first-, second-, and third-order derivatives of the return function. The derivatives of Q, R, and W follow immediately from $\frac{\partial M}{\partial \theta}$, e.g., $\frac{\partial Q}{\partial \theta} = S_x' \frac{\partial M}{\partial \theta} S_x$. The remaining derivations are the same as in the linear-quadratic case.

A linear-quadratic economy with distortions

The optimization problem that we start with is given by

$$\max_{\{\bar{u}_t\}} \sum_{t=0}^{\infty} \beta^t \Big\{ \begin{bmatrix} \bar{y}_t \\ \bar{z}_t \end{bmatrix}' \begin{bmatrix} \bar{Q}_y & \bar{Q}_z \\ \bar{Q}_z' & \bar{Q}_{22} \end{bmatrix} \begin{bmatrix} \bar{y}_t \\ \bar{z}_t \end{bmatrix} + \bar{u}_t' \bar{R} \bar{u}_t + 2 \begin{bmatrix} \bar{y}_t \\ \bar{z}_t \end{bmatrix}' \begin{bmatrix} \bar{W}_y \\ \bar{W}_z \end{bmatrix} \bar{u}_t \Big\} \tag{156}$$

subject to

$$\bar{y}_{t+1} = \bar{A}_y \bar{y}_t + \bar{A}_z \bar{z}_t + \bar{B}_y \bar{u}_t + C\bar{\epsilon}_{t+1}.$$

To ease notation, we convert the problem to one without cross-products or discounting. Let

$$\begin{aligned} y_t &= \beta^{t/2}\bar{y}_t \\ z_t &= \beta^{t/2}\bar{z}_t \\ u_t &= \beta^{t/2}\bar{u}_t \\ \epsilon_t &= \beta^{t/2}\bar{\epsilon}_t \\ R &= \bar{R} \\ Q_y &= \bar{Q}_y - \bar{W}_y \bar{R}^{-1}\bar{W}_y' \\ Q_z &= \bar{Q}_z - \bar{W}_y \bar{R}^{-1}\bar{W}_z' \end{aligned}$$

$$
\begin{aligned}
Q_{22} &= \bar{Q}_{22} - \bar{W}_z \bar{R}^{-1} \bar{W}_z' \\
A_y &= \sqrt{\beta}(\bar{A}_y - \bar{B}_y \bar{R}^{-1} \bar{W}_y') \\
A_z &= \sqrt{\beta}(\bar{A}_z - \bar{B}_y \bar{R}^{-1} \bar{W}_z') \\
B_y &= \sqrt{\beta} \bar{B}_y \\
\Theta &= (I + \bar{\Psi} \bar{R}^{-1} \bar{W}_z')^{-1} (\bar{\Theta} - \bar{\Psi} \bar{R}^{-1} \bar{W}_y') \\
\Psi &= (I + \bar{\Psi} \bar{R}^{-1} \bar{W}_z')^{-1} \bar{\Psi}. \qquad (157)
\end{aligned}
$$

With these definitions, we can restate the optimization problem as follows

$$
\max_{\{u_t\}} \sum_{t=0}^{\infty} \{ \begin{bmatrix} y_t \\ z_t \end{bmatrix}' \begin{bmatrix} Q_y & Q_z \\ Q_z' & Q_{22} \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix}' + u_t' R u_t \} \qquad (158)
$$

subject to

$$
y_{t+1} = A_y y_t + A_z z_t + B_y u_t + C \epsilon_{t+1}.
$$

Let $\hat{A} = A_y + A_z \Theta$, $\hat{Q} = Q_y + Q_z \Theta$, $\hat{B} = B_y + A_z \Psi$, and $\tilde{A} = A_y - B_y R^{-1} \Psi' Q_z'$. The decision function in this case is given by

$$
F = \left(R + B_y' P \hat{B}\right)^{-1} B_y' P \hat{A}, \qquad (159)
$$

where P satisfies

$$
P = \hat{Q} + \tilde{A}' P \hat{A} - \tilde{A}' P \hat{B} (R + B_y' P \hat{B})^{-1} B_y' P \hat{A}. \qquad (160)
$$

The decision function for the original problem is given by

$$
\bar{F} = (\bar{R} + \bar{W}_z' \bar{\Psi})^{-1} (\bar{R} F + \bar{W}_y' + \bar{W}_z' \bar{\Theta}), \qquad (161)
$$

and the equilibrium law of motion for $\bar{y}_t$ is

$$
\bar{y}_{t+1} = A_o \bar{y}_t + C \bar{\epsilon}_{t+1}, \quad A_o = \bar{A}_y + \bar{A}_z \bar{\Theta} - \bar{A}_z \bar{\Psi} \bar{F} - \bar{B}_y \bar{F} = \beta^{-\frac{1}{2}} (\hat{A} - \hat{B} F). \qquad (162)
$$

Therefore, the derivative of A_o with respect to a parameter θ is given by

$$
\frac{\partial A_o}{\partial \theta} = \beta^{-\frac{1}{2}} (\frac{\partial \hat{A}}{\partial \theta} - \frac{\partial \hat{B}}{\partial \theta} F - \hat{B} \frac{\partial F}{\partial \theta}). \qquad (163)
$$

To calculate $\frac{\partial A_o}{\partial \theta}$ requires several steps. First, we need the derivatives of $\hat{A}$, $\hat{B}$, and F with respect to θ:

$$
\frac{\partial \hat{A}}{\partial \theta} = \frac{\partial A_y}{\partial \theta} + \frac{\partial A_z}{\partial \theta} \Theta + A_z \frac{\partial \Theta}{\partial \theta} \qquad (164)
$$

$$
\frac{\partial \hat{B}}{\partial \theta} = \frac{\partial B_y}{\partial \theta} + \frac{\partial A_z}{\partial \theta} \Psi + A_z \frac{\partial \Psi}{\partial \theta} \qquad (165)
$$

$$
\begin{aligned}
\frac{\partial F}{\partial \theta} &= -\big(R + B_y' P B_y + B_y' P A_z \Psi\big)^{-1}\big(\frac{\partial R}{\partial \theta} + \frac{\partial B_y}{\partial \theta}' P B_y \\
&\quad + B_y' \frac{\partial P}{\partial \theta} B_y + B_y' P \frac{\partial B_y}{\partial \theta} + \frac{\partial B_y}{\partial \theta}' P A_z \Psi \\
&\quad + B_y' \frac{\partial P}{\partial \theta} A_z \Psi + B_y' P \frac{\partial A_z}{\partial \theta} \Psi + B_y' P A_z \frac{\partial \Psi}{\partial \theta}\big) F \\
&\quad + \big(R + B_y' P B_y + B_y' P A_z \Psi\big)^{-1}\big(\frac{\partial B_y}{\partial \theta}' P A_y + B_y' \frac{\partial P}{\partial \theta} A_y + B_y' P \frac{\partial A_y}{\partial \theta} \\
&\quad + \frac{\partial B_y}{\partial \theta}' P A_z \Theta + B_y' \frac{\partial P}{\partial \theta} A_z \Theta + B_y' P \frac{\partial A_z}{\partial \theta} \Theta + B_y' P A_z \frac{\partial \Theta}{\partial \theta}\big) \\
&= \big(R + B_y' P \hat{B}\big)^{-1}\big(-\frac{\partial R}{\partial \theta} F + \frac{\partial B_y}{\partial \theta}' P(\hat{A} - \hat{B}F) \\
&\quad + B_y' \frac{\partial P}{\partial \theta}(\hat{A} - \hat{B}F) + B_y' P(\frac{\partial A_y}{\partial \theta} - \frac{\partial B_y}{\partial \theta} F) \\
&\quad + B_y' P \frac{\partial A_z}{\partial \theta}(\Theta - \Psi F) + B_y' P A_z(\frac{\partial \Theta}{\partial \theta} - \frac{\partial \Psi}{\partial \theta} F)\big). && (166)
\end{aligned}
$$

Note that these derivatives are functions of of $\frac{\partial R}{\partial \theta}$, $\frac{\partial B_y}{\partial \theta}$, $\frac{\partial A_y}{\partial \theta}$, $\frac{\partial A_z}{\partial \theta}$, $\frac{\partial \Theta}{\partial \theta}$, $\frac{\partial \Psi}{\partial \theta}$, and $\frac{\partial P}{\partial \theta}$. The derivative of R is given since $R = \bar{R}$. The derivatives for B_y, A_y, A_z, Θ and Ψ follow from their definitions above, e.g.,

$$
\begin{aligned}
\frac{\partial B_y}{\partial \theta} &= \sqrt{\beta}\frac{\partial \bar{B}_y}{\partial \theta}, && (167) \\
\frac{\partial A_y}{\partial \theta} &= \sqrt{\beta}(\frac{\partial \bar{A}_y}{\partial \theta} - \frac{\partial \bar{B}_y}{\partial \theta}\bar{R}^{-1}\bar{W}_y' + \bar{B}_y \bar{R}^{-1}\frac{\partial \bar{R}}{\partial \theta}\bar{R}^{-1}\bar{W}_y' \\
&\quad - \bar{B}_y \bar{R}^{-1}\frac{\partial \bar{W}_y}{\partial \theta}'), && (168) \\
\frac{\partial A_z}{\partial \theta} &= \sqrt{\beta}(\frac{\partial \bar{A}_z}{\partial \theta} - \frac{\partial \bar{B}_y}{\partial \theta}\bar{R}^{-1}\bar{W}_z' + \bar{B}_y \bar{R}^{-1}\frac{\partial \bar{R}}{\partial \theta}\bar{R}^{-1}\bar{W}_z' \\
&\quad - \bar{B}_y \bar{R}^{-1}\frac{\partial \bar{W}_z}{\partial \theta}'), && (169) \\
\frac{\partial \Theta}{\partial \theta} &= -(I + \bar{\Psi}\bar{R}^{-1}\bar{W}_z')^{-1}\big(\frac{\partial \bar{\Psi}}{\partial \theta}\bar{R}^{-1}\bar{W}_z'\Theta - \bar{\Psi}\bar{R}^{-1}\frac{\partial \bar{R}}{\partial \theta}\bar{R}^{-1}\bar{W}_z'\Theta \\
&\quad + \bar{\Psi}\bar{R}^{-1}\frac{\partial \bar{W}_z}{\partial \theta}'\Theta - \frac{\partial \bar{\Theta}}{\partial \theta} + \frac{\partial \bar{\Psi}}{\partial \theta}\bar{R}^{-1}\bar{W}_z' - \bar{\Psi}\bar{R}^{-1}\frac{\partial \bar{R}}{\partial \theta}\bar{R}^{-1}\bar{W}_z' \\
&\quad + \bar{\Psi}\bar{R}^{-1}\frac{\partial \bar{W}_z}{\partial \theta}'\big), && (170) \\
\frac{\partial \Psi}{\partial \theta} &= -(I + \bar{\Psi}\bar{R}^{-1}\bar{W}_z')^{-1}\big(\frac{\partial \bar{\Psi}}{\partial \theta}\bar{R}^{-1}\bar{W}_z'\Psi - \bar{\Psi}\bar{R}^{-1}\frac{\partial \bar{R}}{\partial \theta}\bar{R}^{-1}\bar{W}_z'\Psi \\
&\quad + \bar{\Psi}\bar{R}^{-1}\frac{\partial \bar{W}_z}{\partial \theta}'\Psi - \frac{\partial \bar{\Psi}}{\partial \theta}\big). && (171)
\end{aligned}
$$

The derivative for P is given by

$$\frac{\partial P}{\partial \theta} = \sqrt{\beta}\tilde{A}_o' \frac{\partial P}{\partial \theta} A_o + \frac{\partial \hat{Q}}{\partial \theta} + \sqrt{\beta}\Big[\frac{\partial \tilde{A}}{\partial \theta}' - \tilde{F}' \frac{\partial B_y}{\partial \theta}\Big] P A_o$$
$$+ \tilde{A}_o' P \Big[\frac{\partial \hat{A}}{\partial \theta} - \frac{\partial \hat{B}}{\partial \theta} F\Big] + \tilde{F}' \frac{\partial R}{\partial \theta} F, \tag{172}$$

where $\tilde{F} = (R + B_y' P \hat{B})^{-1} \hat{B} P' \tilde{A}$, $\tilde{A}_o = \tilde{A} - B_y \tilde{F}$, and

$$\frac{\partial \hat{Q}}{\partial \theta} = \frac{\partial Q_y}{\partial \theta} + \frac{\partial Q_z}{\partial \theta} \Theta + Q_z \frac{\partial \Theta}{\partial \theta} \tag{173}$$

$$\frac{\partial \tilde{A}}{\partial \theta} = \frac{\partial A_y}{\partial \theta} - \frac{\partial B_y}{\partial \theta} R^{-1} \Psi' Q_z' + B_y R^{-1} \frac{\partial R}{\partial \theta} R^{-1} \Psi' Q_z'$$
$$- B_y R^{-1} \frac{\partial \Psi}{\partial \theta}' Q_z' - B_y R^{-1} \Psi' \frac{\partial Q_z}{\partial \theta}'. \tag{174}$$

The last two derivatives needed are $\frac{\partial Q_y}{\partial \theta}$ and $\frac{\partial Q_z}{\partial \theta}$:

$$\frac{\partial Q_y}{\partial \theta} = \frac{\partial \bar{Q}_y}{\partial \theta} - \frac{\partial \bar{W}_y}{\partial \theta} \bar{R}^{-1} \bar{W}_y' + \bar{W}_y \bar{R}^{-1} \frac{\partial \bar{R}}{\partial \theta} \bar{R}^{-1} \bar{W}_y' - \bar{W}_y \bar{R}^{-1} \frac{\partial \bar{W}_y}{\partial \theta}', \tag{175}$$

$$\frac{\partial Q_z}{\partial \theta} = \frac{\partial \bar{Q}_z}{\partial \theta} - \frac{\partial \bar{W}_y}{\partial \theta} \bar{R}^{-1} \bar{W}_z' + \bar{W}_y \bar{R}^{-1} \frac{\partial \bar{R}}{\partial \theta} \bar{R}^{-1} \bar{W}_z' - \bar{W}_y \bar{R}^{-1} \frac{\partial \bar{W}_z}{\partial \theta}'. \tag{176}$$

We now have everything that we need to compute the derivatives of the matrices in the decision rule and the law of motion for the state vector. To avoid iterating on Eq. (172) for $\frac{\partial P}{\partial \theta}$, we instead take the gradient, e.g.,

$$\nabla_\theta P = (I - \sqrt{\beta} A_o' \otimes \tilde{A}_o')^{-1} \{ \nabla_\theta \hat{Q} + (I \otimes \tilde{A}_o' P') \nabla_\theta \hat{A}$$
$$+ \sqrt{\beta}(A_o' P' \otimes I) \nabla_\theta \tilde{A}' - (F' \otimes \tilde{A}_o' P') \nabla_\theta \hat{B}$$
$$- \sqrt{\beta}(A_o' P' \otimes \tilde{F}') \nabla_\theta B_y' + (F' \otimes \tilde{F}') \nabla_\theta R. \tag{177}$$

Thus, the gradient of F is given by

$$\nabla_\theta F = (I \otimes \mathcal{R} B_y' P) \nabla_\theta A_y + \big((\Theta - \Psi F') \otimes \mathcal{R} B_y' P\big) \nabla_\theta A_z$$
$$- (F' \otimes \mathcal{R} B_y' P) \nabla_\theta B_y + \sqrt{\beta}(A_o' P' \otimes \mathcal{R}) \nabla_\theta B_y'$$
$$+ \sqrt{\beta}(A_o' \otimes \mathcal{R} B_y') \nabla_\theta P - (F' \otimes \mathcal{R}') \nabla_\theta R$$
$$+ (I \otimes \mathcal{R} B_y' P A_z) \nabla_\theta \Theta - (F' \otimes \mathcal{R} B_y' P A_z) \nabla_\theta \Psi, \tag{178}$$

where $\mathcal{R} = (R + B_y' P \hat{B})^{-1}$. In terms of the computer code, we start with Eqs. (167)-(171) and Eqs. (175)-(176), which relate the derivatives of the original problem to those of the problem without discounting or cross-product terms. To compute the gradients of these objects in terms of our inputs, we use the fact that $vec(ABC) = (C' \otimes A)vec(B)$ for any matrices A, B, and C with the appropriate dimensions such that ABC exists. We next compute the derivatives for $\hat{A}$, $\hat{B}$, $\hat{Q}$, and $\tilde{A}$ which appear in Eqs. (164), (165), (173), and (174). Finally, we compute $\nabla_\theta P$ in Eq. (177), $\nabla_\theta F$ in Eq. (178), and

$$\nabla_\theta A_o = \beta^{-\frac{1}{2}}(\nabla_\theta \hat{A} - (F' \otimes I)\nabla_\theta \hat{B} - (I \otimes \hat{B})\nabla_\theta F).$$

A nonlinear economy with distortions

The optimization problem that we start with is given by

$$\begin{aligned} &\max_{\{u_t\}} E_0 \sum_{t=0}^{\infty} \beta^t r(Z_t, \theta) \\ &\quad \text{subject to } \bar{y}_{t+1} = \bar{A}_y \bar{y}_t + \bar{A}_z \bar{z}_t + \bar{B}_y \bar{u}_t + C\bar{\epsilon}_{t+1} \\ &\qquad\qquad\qquad Z_t = [\bar{y}_t', \bar{z}_t', \bar{u}_t']'. \end{aligned}$$

As in the case of the economy without distortions, we solve a related problem that has the same form as the problem of Eq. (156). The approximation method is the same as in the model without distortions; thus, all of the required derivatives have already been computed.

References

Anderson, Brian D.O. and John B. Moore (1979). *Optimal Filtering*. Englewood Cliffs: Prentice-Hall.

Blanchard, Oliver J. and Charles M. Kahn (1980). 'The Solution of Linear Difference Equations Under Rational Expectations'. *Econometrica*, Vol. 48, pp. 1305–1311.

Dagli, C. and J. Taylor (1984). 'Estimation and Solution of Linear Rational Expectations Models Using a Polynomial Matrix Factorization'. *Journal of Economic Dynamics and Control*, Vol. 8, No. 3, pp. 341–348.

Doan, Thomas, Robert Litterman, and Christopher Sims (1984). 'Forecasting and Conditional Projection Using Realistic Prior Distributions'. *Econometric Reviews*, Vol. 3, No. 1, pp. 1–100.

Fletcher, R. (1987). *Practical Methods of Optimization*. New York: John Wiley & Sons.

Hansen, L.P. and T.J. Sargent (1994). 'Recursive Linear Models of Dynamic Economies'. Mimeo. University of Chicago.

İmrohoroğlu, Selahattin (1993). 'Testing for Sunspot Equilibria in the German Hyperinflation'. *Journal of Economic Dynamics and Control*, Vol. 17, No. 1/2, pp. 289–318.

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo (1988a). 'Production, Growth, and Business Cycles: I. The Basic Neoclassical Model'. *Journal of Money, Credit, and Banking*, Vol. 21, No. 2/3, pp. 195–232.

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo (1988b). 'Production, Growth, and Business Cycles: II. New Directions'. *Journal of Money, Credit, and Banking*, Vol. 21, No. 2/3, pp. 309–342.

Kwakernaak, H. and R. Sivan (1972). *Linear Optimal Control Systems*. New York: Wiley Interscience.

Kydland, Finn and Edward C. Prescott (1982). 'Time to Build and Aggregate Fluctuations'. *Econometrica*, Vol. 50, pp. 1345–1370.

McGrattan, Ellen R. (1994). 'A Note on Computing Competitive Equilibria in Linear Models'. *Journal of Economic Dynamics and Control*, Vol. 18, pp. 149–160.

McGrattan, Ellen, Richard Rogerson, and Randall Wright (1993). 'Household Production and Taxation in the Stochastic Growth Model'. Mimeo. Staff Report 166, Federal Reserve Bank of Minneapolis.

Sargent, T.J. (1980). 'Notes on Linear Filtering, Control, and Rational Expectations Models'. Mimeo. Federal Reserve Bank of Minneapolis.

Rosen, Sherwin, Kevin M. Murphy, and José A. Scheinkman (1994). 'Cattle Cycles'. *Journal of Political Economy*, Vol. 102, No. 3, pp. 468–492.

Sims, C.A. (1980). 'Macroeconomics and Reality'. *Econometrica*, Vol. 48, No. 1, pp. 1–48.

United States, Bureau of the Census (1975). *Historical Statistics of the United States, Colonial Times to 1970*. Washington, D.C.: U.S. Department of Commerce.

United States, Bureau of the Census (1989). *Agricultural Statistics*. Washington, D.C.: U.S. Department of Commerce.

Vaughan, D.R. (1969). 'A Negative-Exponential Solution to the Matrix Riccati Equation'. *IEEE Transactions on Automatic Control*, Vol. 14, pp. 72–75.

Vaughan, D.R. (1970). 'A Nonrecursive Algebraic Solution for the Discrete Riccati Equation'. *IEEE Transactions on Automatic Control*, Vol. AC-15, No. 5, October, pp. 597–599.

Whiteman, Charles (1983). *Linear Rational Expectations Models: A User's Guide*. Minneapolis: University of Minnesota Press.

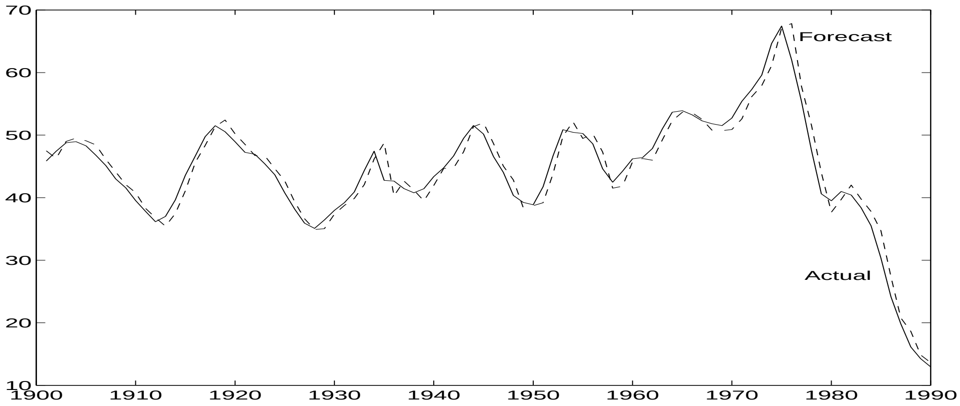

Figure 1. One-step-ahead forecast and actual total stock.

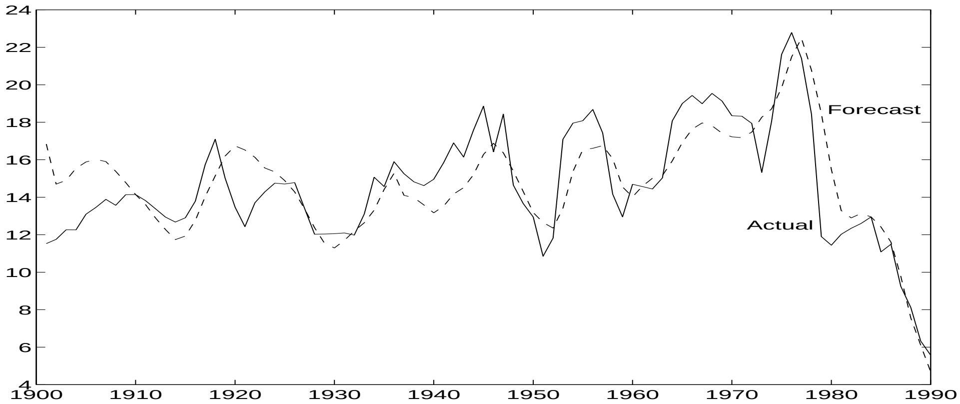

Figure 2. One-step-ahead forecast and actual of slaughter rate.

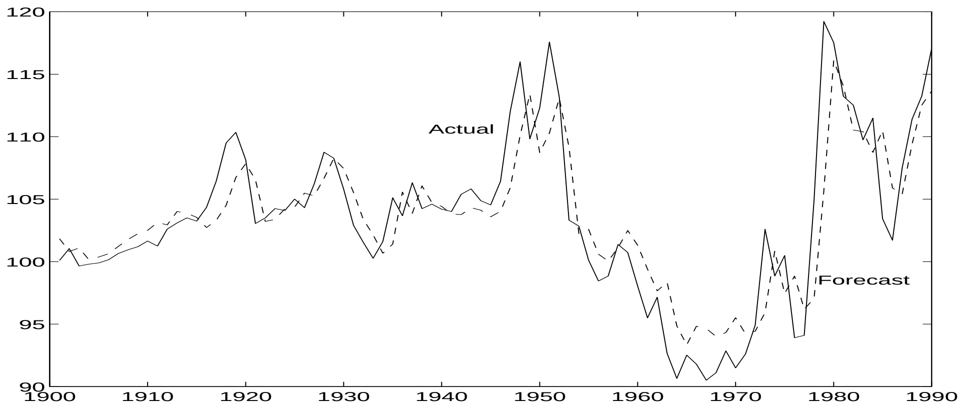

Figure 3. One-step-ahead forecast and actual price of slaughtered beef.