

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 728

June 2002

FINDING NUMERICAL RESULTS TO LARGE SCALE ECONOMIC
MODELS USING PATH-FOLLOWING ALGORITHMS:
A VINTAGE CAPITAL EXAMPLE

Brett D. Berger

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDP's are available on the Web at www.federalreserve.gov/pubs/ifdp/

FINDING NUMERICAL RESULTS TO LARGE SCALE ECONOMIC
MODELS USING PATH-FOLLOWING ALGORITHMS:
A VINTAGE CAPITAL EXAMPLE

Brett D. Berger*

Abstract: This paper describes the numerical optimization methods used in Berger (2001) to find the complete time paths of key economic variables in neoclassical vintage capital models. An interior and a non-interior point method are discussed. Both of the methods are part of the general class of “path-following” algorithms. These algorithms can be efficiently applied to convex programming problems; and due to the standard shape of production and utility functions, many economic problems can be written as convex programming problems. Vintage capital models add scale and complexity to standard growth models because one must now handle the dynamics of multiple capital stocks. This increase in complexity will often prevent the discovery (or existence) of closed form solutions, making numerical solutions of the type found in Berger (2001) necessary.

Keywords: optimization, productivity, technology

*The author is an economist in the Division of International Finance, Board of Governors of the Federal Reserve System. He can be reached at Mail Stop 42B, Federal Reserve Board, Washington, DC 20551; brett.d.berger@frb.gov. The majority of this research was done while the author was a graduate student at the University of Washington. The author is grateful for the comments from his supervisory committee, Stephen Turnovsky, Charles Engel, Richard Hartman, and Terry Rockafellar, and for comments from colleagues at the Board. The views of this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System.

1. Introduction

To more accurately mirror the world, economic models have become increasingly complex. In many cases, it is necessary to find numerical solutions to parameterized versions of the models if one is to obtain information on transitional dynamics. Optimization versions of vintage capital models are an example of such models.

There are two general methods for modeling technology, “embodied” and “disembodied.” Vintage capital or “embodied technology” models are models in which technology is inherent in the capital stock of the economy (i.e. the physical equipment, buildings, etc.). The embodiment of technology implies a heterogeneous capital stock, and that it is necessary to invest in new capital if the economy is to reap the benefits of new technology. It is obvious that much of the technological growth in the world is of the embodied variety, however, most economic models assume technology is disembodied owing to the difficulty of tracking multiple capital stocks.

This paper discusses numerical optimization methods that are capable of handling the large-scale nature of vintage capital models, and their application to the three basic neoclassical vintage capital models: putty-putty, clay-clay, and putty-clay.¹ The optimization methods used in this paper have the potential for application to a wide array of economic models. Diminishing returns in both productive factors and utility often implies that models can be posed as convex programming problems, which mathematical theory states will have global solutions (assuming a solution exists). The algorithms

¹ A detailed description of the results can be found in Berger (2001).

discussed in this paper are guaranteed to converge when applied to convex programming problems.

2. Optimization Method 1: Interior Point Method

The algorithms used to obtain the numerical solutions of Berger (2001) are part of the general class of “path-following” algorithms. An interior point method was utilized to obtain approximations to the solutions of the putty-putty and clay-clay models.

Interior point methods are so named because the choice variables lie in the strictly feasible set of the optimization problem, (P) , at each iteration. This is achieved by the addition of penalty functions, f_1 , to the objective function, f_0 . The penalty functions are functions of the constraint values. The penalty functions increase in value as the boundary of the feasible set is approached, and take on infinite values at the boundary, thereby creating a “barrier” to exiting the feasible set. The penalty functions are multiplied by a positive scalar, μ_k , which is monotonically decreasing in the iterations, k . The solutions to this barrier problem, $(P(\mu_k))$, as $\mu_k \rightarrow 0$, comprise the “central path” which is followed to the solution of (P) . Below is the formal statement of the standard problem (P) and the resulting optimality conditions.

$$\begin{aligned}
& \text{Minimize } f_0(x) \\
& \text{over } x \in R^n \\
& \text{subject to:} \\
& \quad G(x) \leq 0 \\
& \text{where:} \\
& \quad f_0 : R^n \mapsto R \\
& \quad G : R^n \mapsto R^m \\
& \quad G(x) = (g_1(x), g_2(x), \dots, g_m(x))' \quad \text{with } g_i : R^n \mapsto R
\end{aligned} \tag{P}$$

(P) is a general problem formulation because equality constraints can be included by splitting the constraint into two inequality constraints. The lagrangian for (P) is:

$$L(x; \lambda) = f_0(x) + \lambda' G(x) \tag{1}$$

This leads to the optimality conditions of (P):

$$\nabla f_0 + \nabla G(x)' \lambda = 0 \tag{2}$$

$$G(x) \leq 0 \tag{3}$$

$$\lambda \geq 0 \tag{4}$$

$$\Lambda G(x) = 0 \tag{5}$$

where Λ is a diagonal matrix with λ on the diagonal. (5) is the complementarity condition stating that either λ_i , g_i , or both must equal zero for all i . Let the penalty function be the standard log barrier function, such that for $z \in R^m$:

$$\ln(z) = \begin{cases} \sum_{i=1}^m \ln(z_i) & \text{if } z_i > 0 \quad \forall i \\ \infty & \text{otherwise} \end{cases} \tag{6}$$

Thus, $(P(\mu_k))$, the log barrier version of the standard problem, is:

$$\begin{aligned} &\text{Minimize } f_0(x) - \mu_k \ln(-G(x)) \\ &\text{over } x \in R^n \end{aligned} \quad (P(\mu_k))$$

where μ_k is a positive scalar. This leads to the optimality conditions:

$$\nabla f_0 + \nabla G(x)' \overline{\mu_k - G(x)}^{-1} e = 0 \quad (7)$$

$$G(x) < 0 \quad (8)$$

where $\overline{z} = \text{diag}(z) \in R^{m \times m}$ and $e = (1, \dots, 1)' \in R^m$. (8) must hold at a solution because otherwise the value of the objective is infinity by the definition of the penalty function.

Convex programming problems are minimization problems in which the objective function and the inequality constraints (written as $g_i \leq 0$) are convex functions, and the equality constraints (written as $g_i = 0$) are affine functions. Theory states that all solutions to convex programming problems are global solutions. Further, when a solution exists, convex programming problems with strictly convex objective functions have a unique, global solution.

For convex programming problems, Proposition 4.1.1 of Bertsekas (1999) states that “every limit point of a sequence $\{x_k\}$ generated by a barrier method is a global minimum of the original constrained problem.” In other words, the sequence $\{x_k\}$, the solutions to the problems $(P(\mu_k))$ as $\mu_k \rightarrow 0$, goes to a global minimum, \bar{x} , of (P) .

Both (P) and $(P(\mu_k))$ can be rewritten in equivalent equilibrium formats where slack variables, restricted to be non-negative, are added to the inequality constraints to make them equality constraints. This is advantageous because of the resulting structure of the Jacobian of the optimality conditions, and readily obtained dual variables. By two

problems being “equivalent” it is meant that a stationary point of one problem is also a stationary point of the other.

The equilibrium formatted problem, (E) and its corresponding lagrangian follow:

$$\begin{aligned}
 &\text{Minimize } f_0(x) \\
 &\text{over } x \in R^n \\
 &\text{subject to:} \\
 &\quad s \geq 0 \\
 &\quad G(x) + s = 0
 \end{aligned} \tag{E}$$

$$L(x, s; \lambda) = f_0(x) + \lambda'(G(x) + s) + \gamma'(-s) \tag{9}$$

(9) leads to the optimality conditions of (E) :

$$\nabla f_0 + \nabla G(x)' \lambda = 0 \tag{10}$$

$$\lambda - \gamma = 0 \quad \Rightarrow \quad \lambda = \gamma \tag{11}$$

$$G(x) + s = 0 \tag{12}$$

$$-s \leq 0 \tag{13}$$

$$\gamma \geq 0 \tag{14}$$

$$\Gamma s = 0 \quad \text{where} \quad \Gamma = \text{diag}(\lambda) \tag{15}$$

It is simple to show that the optimality conditions of (P) are equivalent to those of (E) .

$(E(\mu_k))$, is formed by using the log barrier function to force s to lie in the positive orthant.

$$\begin{aligned}
 &\text{Minimize } f_0(x) - \mu_k \ln(s) \\
 &\text{over } x \in R^n \quad s \in R^m \\
 &\text{subject to:} \\
 &\quad G(x) + s = 0
 \end{aligned} \tag{E(\mu_k)}$$

It is not necessary to have an explicit non-negativity constraint for s since that will necessarily be the case at a solution. The lagrangian for $(E(\mu_k))$ is given by (16) and followed by the corresponding optimality conditions:

$$L(x, s; \lambda) = f_0(x) - \mu_k \ln(s) + \lambda'(G(x) + s) \quad (16)$$

$$\nabla f_0 + \nabla G(x)' \lambda = 0 \quad (17)$$

$$-\mu_k S^{-1} e + \lambda = 0 \quad \Rightarrow \quad S \lambda = \mu_k e \quad (18)$$

$$G(x) + s = 0 \quad (19)$$

$$s \geq 0 \quad (20)$$

where $S = \text{diag}(s)$. (18) and (19) imply that $s = -G(x)$ and $\lambda = \mu_k \left(\overline{-G(x)}^{-1} \right) e$. Then

(7) is equivalent to (17); and (8) is equivalent to (19) combined with (20). Therefore

$(P(\mu_k))$ and $(E(\mu_k))$ are equivalent problems. Hence, one can work with the

equilibrium form of the problem and be assured that the theoretical results corresponding to convex programming and barrier problems will apply.

Most applications of the interior-point algorithm will not strictly follow the central path to the solution but instead reduce μ_k by a variable amount at each iteration. This was true in obtaining the results of Berger (2001). For that paper, the variables also were not restricted to lie in the strictly feasible region of (P) for all iterations. Instead, using formulation $(E(\mu_k))$, only the restrictions that x , s , and λ are positive are enforced at each iteration. $(G(x) + s) \rightarrow 0$ with $s > 0$, but the solution at termination must be checked to make sure that $G(x) \leq 0$ since s can converge to zero along with $G(x)$ and

therefore the solution at termination of the algorithm could lie just outside the boundary of feasible set (by the value of the tolerance for the optimality conditions). For many applications, the level of this discrepancy is of little practical value (such as in the vintage capital case).

The iterations use the Newton method and it remains to show that the method is defined for each iteration. We will do this by examining the Jacobian of the optimality conditions of $(E(\mu_k))$. In certain cases it is preferable, for numerical stability reasons, to multiply (17) by a positive definite diagonal matrix D . Without loss of generality (i.e. $D=I$), (17) can be rewritten as:

$$D(\nabla f_0 + \nabla G(x)' \lambda) = 0 \quad (21)$$

The Jacobian of the optimality functions can be written as:

$$J = \begin{bmatrix} M_1 & M_2 & 0 \\ M_3 & 0 & I \\ 0 & S & \Lambda \end{bmatrix} \quad (22)$$

where

$$M_1 = D(\nabla^2 f_0(x) + \nabla^2 G(x)' \lambda), \quad M_1 \in R^{n \times n} \quad (23)$$

$$M_2 = D(\nabla G(x)'), \quad M_2 \in R^{n \times m} \quad (24)$$

$$M_3 = \nabla G(x), \quad M_3 \in R^{m \times n} \quad (25)$$

with

$$\nabla^2 G(x)' \lambda = \lambda_1 \nabla^2 g_1(x) + \lambda_2 \nabla^2 g_2(x) + \dots + \lambda_m \nabla^2 g_m(x)$$

The Newton direction is $d_k = -J_k^{-1} g_k$ where g_k , the vector of optimality functions, and J_k are both evaluated at the k^{th} iteration values of x , s , and λ . From the

definition of d_k , there is a unique Newton direction for each iteration if J is nonsingular.

J is nonsingular if only if there is a unique solution to the following system:

$$\begin{bmatrix} M_1 & M_2 & 0 \\ M_3 & 0 & I \\ 0 & S & \Lambda \end{bmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad (26)$$

From (26), v and w can be solved in terms of u by the following method:

$$w = b - M_3 u \quad (27)$$

$$v = S^{-1}(c - \Lambda w) \quad (28)$$

Substituting (27) and (28) into (26):

$$(M_2 S^{-1} \Lambda M_3 + M_1) u = a - M_2 S^{-1}(c - \Lambda b) \quad (29)$$

From (29), the Newton step has a unique solution if and only if the matrix

$(M_2 S^{-1} \Lambda M_3 + M_1)$ is nonsingular. It will be shown that under certain conditions

$(M_2 S^{-1} \Lambda M_3 + M_1)$ is positive definite and therefore nonsingular.

Proposition 1. *Let $x \in R_+^n$ (non-negative orthant) and (P) be a convex programming problem. Then the Newton direction has a unique solution.*

Proof. As shown above, the Newton direction has a unique solution if and only if

$(M_2 S^{-1} \Lambda M_3 + M_1)$ is non-singular. Suppose $x \in R_+^n$ and (P) is a convex programming problem.

First will be shown that $M_2 S^{-1} \Lambda M_3$ is positive definite. Then it will be shown that M_1 is positive semi-definite. The sum of a positive definite and positive semi-

definite matrix is positive definite and non-singular. Hence, this is sufficient for the proof.

From (24) and (25), $M_2 S^{-1} \Lambda M_3 = D M_3' S^{-1} \Lambda M_3 = D H' H$, where D is positive definite and $H = (S^{-1} \Lambda)^{1/2} M_3 = (S^{-1} \Lambda)^{1/2} \nabla G(x)$.

With no abstract constraints in (P) , $x \in R_+^n$ implies that $\nabla G(x) \in R^{m \times n}$ has rank n , with $m \geq n$. Then $\text{nul}(G(x)) = 0$. Since at each iteration S^{-1} and Λ are positive definite diagonal matrices, this implies that $\text{nul}(H) = 0$. It is obvious that $(Hw)' Hw \geq 0$, and $\text{nul}(H) = 0$ implies that $(Hw)' Hw = 0$ if and only if $w = 0$. Let $Q = H' H$. Then $w' H' Hw = w' Qw$. Since $w' Qw > 0 \quad \forall w \neq 0$, Q is positive definite. Since D is a diagonal positive definite matrix, $DQ = D H' H = D M_3' S^{-1} \Lambda M_3 = M_2 S^{-1} \Lambda M_3$ is positive definite.

$M_1 = D(\nabla^2 f_0(x) + \nabla^2 G(x)' \lambda)$. So, M_1 is positive semi-definite if and only if $\nabla^2 f_0(x) + \nabla^2 G(x)' \lambda$ is positive semi-definite. If (P) is a convex programming problem then $\nabla^2 f_0(x)$, the Hessian of the objective function, is positive semi-definite. Similarly, $\nabla^2 G(x)' \lambda = \lambda_1 \nabla^2 g_1(x) + \lambda_2 \nabla^2 g_2(x) + \dots + \lambda_m \nabla^2 g_m(x)$, and if (P) is a convex programming problem, then each g_i is convex and $\nabla^2 g_i$ is positive semi-definite. Therefore, by the fact that a sum of positively weighted positive semi-definite matrices is positive semi-definite, $\nabla^2 G(x)' \lambda$ is positive semi-definite. By the same fact,

$\nabla^2 f_0(x) + \nabla^2 G(x)' \lambda$ is positive semi-definite. Therefore, since D is a diagonal positive definite matrix, M_1 is positive semi-definite. □

The method of reducing μ_k in Berger (2001) was adapted from Wright (1997). At each iteration, μ_k is defined to be the average value of the complementarity conditions (i.e. $\mu_k = (\lambda' s) / m$). The initial guess for λ and s give μ_0 . A parameter $\sigma \in (0, 1)$ determines by what fraction the next iteration should attempt to decrease μ_k ($\mu_{k+1} = \sigma \mu_k$). There will be a speed tradeoff in the choice of σ . A lower value of σ may allow for fewer iterations if at each iteration the step stays close to the central path. However, too small a value of σ may cause the step to diverge far enough from the central path that the Newton method does not converge well on subsequent iterations.

3. Non-interior Path Following Algorithm

Solutions to the putty-clay model in Berger (2001) were obtained using a non-interior path-following algorithm based on Burke and Xu (2000). Burke and Xu, however, present a non-interior path following algorithm for a linear complementarity problem, whereas the putty-clay model is a general non-linear problem. As such a problem, multiple local solutions are possible.

The following is the format of a general non-linear optimization problem (except for the abstract constraint on \tilde{x}):

Minimize $f_0(\tilde{x})$

over $\tilde{x} \in R_+^n$ (non-negative orthant)

subject to:

$$G(\tilde{x}) \leq 0$$

where:

(P_{gnl})

$$f_0 : R^n \mapsto R$$

$$G : R^n \mapsto R^m$$

$$G(\tilde{x}) = (g_1(\tilde{x}), g_2(\tilde{x}), \dots, g_m(\tilde{x}))' \quad \text{with } g_i : R^n \mapsto R$$

The following is the lagrangian corresponding to (P_{gnl}) :

$$L(\tilde{x}; \lambda) = f_0(\tilde{x}) + \lambda' G(\tilde{x}) \quad (30)$$

The first order optimality conditions are:

$$\frac{\partial L}{\partial \tilde{x}} = \nabla f_0(\tilde{x}) + \nabla G(\tilde{x})' \lambda \geq 0 \quad (31)$$

$$\frac{\partial L}{\partial \lambda} = G(\tilde{x}) \leq 0 \quad (32)$$

Let $x = (\tilde{x}, \lambda)'$. Let $M(x) = \left(\frac{\partial L}{\partial \tilde{x}}, -\frac{\partial L}{\partial \lambda} \right)$. Then the Karush-Kuhn-Tucker (KKT)

conditions are equivalent to and are satisfied for an (x, v) combination of variables if:

$$M(x) - v = 0 \quad (33)$$

$$x \geq 0 \quad (34)$$

$$v \geq 0 \quad (35)$$

$$Vx = 0 \quad (36)$$

where $V = \text{diag}(v)$.

For non-interior path-following algorithms, penalty functions are not used.

Instead, the scalar $\mu_k \geq 0$ is a parameter in a function $\phi(x, v, \mu_k) : R^{(2n+2m+1)} \mapsto R^{(n+m)}$,

where $\phi(x, v, \mu_k) = [\phi(x_1, v_1, \mu_k), \dots, \phi(x_{n+m}, v_{n+m}, \mu_k)]$, $v = M(x)$ and:

$$\phi(x_i, v_i, \mu_k) = x_i + v_i - \sqrt{(x_i - v_i)^2 + 4\mu_k^2} \quad (37)$$

The optimality conditions of (P_{gnl}) will hold if and only if $\phi = 0$ and $\mu_k = 0$. The central path is defined as the (x, v) combinations such that $\phi = 0$ for a sequence of non-negative μ_k . The Newton method is once again used to find a stationary point of the optimization problem. Reductions in μ_k are done in a predictor-corrector step framework. A predictor step, which attempts a reduction of μ_k to zero, is first used in the iteration. A corrector step is then utilized in which only a partial reduction in μ_k is attempted. It is called a ‘‘corrector’’ step because it is used to guide values closer to the central path.

Note 1. $\phi(x_i, v_i, \mu_k) = 0$ if and only if $x_i \geq 0$, $v_i \geq 0$ and $x_i v_i = \mu_k^2$

Proof. Let $\phi(x_i, v_i, \mu_k) = 0$.

Then $x_i + v_i = \sqrt{(x_i - v_i)^2 + 4\mu_k^2}$.

Squaring both sides: $x_i^2 + v_i^2 + 2x_i v_i = (x_i - v_i)^2 + 4\mu_k^2 \Rightarrow 4x_i v_i = 4\mu_k^2 \Rightarrow$

$$x_i v_i = \mu_k^2 \quad (38)$$

By (38), $\mu_k \neq 0$ implies x_i and v_i are non-zero and the same sign.

But $x_i, v_i < 0$ implies that $\phi(x_i, v_i, \mu_k) < 0$. Therefore:

$$\mu_k \neq 0 \quad \Rightarrow \quad x_i, v_i > 0 \quad (39)$$

By (38), $\mu_k = 0$ implies x_i and/or v_i equals zero. By (37), if $\phi = 0$ and either x_i or v_i is zero, then the other is non-negative. Therefore $x \geq 0$, $v \geq 0$ and $xv = \mu_k^2$. It is trivial to see that if $x_i v_i = \mu_k^2$, then $\phi(x_i, v_i, \mu_k) = 0$. \square

Let:

$$F(x, v, \mu_k) = \begin{bmatrix} M(x) - v \\ \phi(x, v, \mu_k) \\ \mu_k \end{bmatrix} \quad \text{with} \quad \phi(x, v, \mu_k) = \begin{bmatrix} \phi(x_1, v_1, \mu_k) \\ \vdots \\ \phi(x_n, v_n, \mu_k) \end{bmatrix} \quad (40)$$

Note 1 shows that (33)-(36) hold if and only if $F = 0$. Newton's method leads to solving (41) for the predictor step:

$$F(x, v, \mu_k) + \nabla F(x, v, \mu_k) \begin{bmatrix} \Delta x \\ \Delta v \\ \Delta \mu_k \end{bmatrix} = 0 \quad (41)$$

where:

$$\nabla F(x, v, \mu_k) = \begin{bmatrix} \nabla M(x) & -I & 0 \\ \nabla_x \phi(x, v, \mu_k) & \nabla_v \phi(x, v, \mu_k) & \nabla_{\mu_k} \phi(x, v, \mu_k) \\ 0 & 0 & 1 \end{bmatrix}$$

Then solve the following for the corrector step:

$$F(x, v, \mu_k) + \nabla F(x, v, \mu_k) \begin{bmatrix} \Delta x \\ \Delta v \\ \Delta \mu_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ (1 - \sigma)\mu_k \end{bmatrix} \quad (42)$$

Proposition 2. *The Newton direction has a unique solution if and only if*

$\nabla_x \phi(x, v, \mu_k) + \nabla_v \phi(x, v, \mu_k) \nabla M(x)$ *is nonsingular. This will occur if* $\nabla M(x)$ *is positive*

semi-definite. $\nabla M(x)$ is positive semi-definite if (P_{gnl}) is a convex programming problem.

Proof. The Newton direction has a unique solution if and only if there is a unique solution to the following system of equations:

$$\begin{bmatrix} \nabla M(x) & -I & 0 \\ \nabla_x \phi(x, v, \mu_k) & \nabla_v \phi(x, v, \mu_k) & \nabla_{\mu_k} \phi(x, v, \mu_k) \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad (43)$$

By (43):

$$w = c \quad (44)$$

$$v = \nabla M u - a \quad (45)$$

By (43), (44) and (45):

$$(\nabla_x \phi + \nabla_v \phi \nabla M) u = b + \nabla_v \phi a - \nabla_{\mu_k} \phi c \quad (46)$$

Hence, there is a unique Newton direction if $(\nabla_x \phi + \nabla_v \phi \nabla M)$ is non-singular.

$$\nabla_{x_i} \phi(x_i, v_i, \mu_k) = 1 - \frac{x_i - v_i}{\sqrt{(x_i - v_i)^2 + 4\mu_k^2}} \quad (47)$$

$$\nabla_{v_i} \phi(x_i, v_i, \mu_k) = 1 + \frac{x_i - v_i}{\sqrt{(x_i - v_i)^2 + 4\mu_k^2}} \quad (48)$$

$$\nabla_x \phi(x, v, \mu_k) = \text{diag}(\nabla_{x_1} \phi(x_1, v_1, \mu_k), \dots, \nabla_{x_n} \phi(x_n, v_n, \mu_k)) \quad (49)$$

$$\nabla_v \phi(x, v, \mu_k) = \text{diag}(\nabla_{v_1} \phi(x_1, v_1, \mu_k), \dots, \nabla_{v_n} \phi(x_n, v_n, \mu_k)) \quad (50)$$

It can be seen from (47)-(50) that $\mu_k > 0$ implies:

$$\nabla_{x_i} \phi(x_i, v_i, \mu_k) > 0 \quad \text{and} \quad \nabla_x \phi(x, v, \mu_k) \text{ is positive definite} \quad (51)$$

$$\nabla_{v_i} \phi(x_i, v_i, \mu_k) > 0 \quad \text{and} \quad \nabla_v \phi(x, v, \mu_k) \text{ is positive definite} \quad (52)$$

Therefore, $(\nabla_x \phi + \nabla_v \phi \nabla M)$ is non-singular if ∇M is positive semi-definite.

$$\nabla M = \begin{bmatrix} M_1 & M_2 \\ -M_2' & 0 \end{bmatrix} \quad (53)$$

where:

$$M_1 = \nabla^2 f_0(x) + \nabla^2 G(x)' \lambda$$

$$M_2 = \nabla G(x)'$$

However:

$$[u \quad v] \nabla M \begin{bmatrix} u \\ v \end{bmatrix} = u' M_1 u \quad (54)$$

By (54), ∇M is positive semi-definite if and only if M_1 is positive semi-definite.

M_1 is positive semi-definite if (P_{gnl}) is a convex programming problem (see proof of Proposition 1). □

4. General Model Descriptions

Berger (2001) found solutions to discrete time, finite horizon, parameterized, optimization versions of the three main types of vintage capital models. These models are referred to as putty-putty, clay-clay, and putty-clay models and vary according to the substitutability of factors. Putty refers to the ability to pair labor and capital in any production ratio. Clay refers to a fixed capital-labor ratio. The first word in each pair refers to substitutability at the time of installation of the capital; and the second word refers to the substitutability for all time after the installation. For example, the term

putty-clay refers to a model in which the capital-labor ratio for capital of a particular vintage may be chosen at the time of installation, but for all time thereafter labor must be used with that vintage of capital according to the chosen proportion.

Each of the three models is based on the basic consumption-savings model. A central planner of the economy chooses a time path of consumption with the objective of maximizing the sum of discounted utility of the population, which is solely a function of consumption. Utility is represented by a finite, constant-elasticity-of-intertemporal-substitution utility function:

$$\begin{aligned} U(C_t) &= \frac{1}{1-\gamma} (C_t^{1-\gamma} - 1) & \gamma \neq 1 \\ &= \ln(C_t) & \gamma = 1 \end{aligned} \tag{55}$$

The elasticity of intertemporal substitution equals $-\frac{1}{\gamma}$, and for fixed C_t this is a continuous function of γ .

Output not consumed each period is invested and becomes capital that can be used to increase production in future periods. Each period in the putty-putty and clay-clay models a static optimization takes place in which the central planner maximizes total output given the capital stock, labor stock, and given parameters. For non-vintage models this allocation is trivial. However, for vintage capital models this is a non-trivial problem in which scarce labor must be assigned to the different vintages of capital.

The three models are based on a Cobb-Douglas production function:

$$f_v(N_v, K_v) = d_t A_v N_v^{1-\alpha} K_v^\alpha \tag{56}$$

where v represents the vintage, d_t is the disembodied technology parameter, A_v is the embodied technology parameter, and $0 < \alpha < 1$. Total output for period t is:

$$f_t = d_t \sum_{v=1}^{V+t-1} A_v N_v^{1-\alpha} K_v^\alpha \quad (57)$$

where V is the number of vintages available in the initial period.. One advantage of the putty-putty model with this production function is that the static optimization problem of maximizing output giving existing resources can be written as a convex programming problem with a strictly convex objective. Therefore, since the problem has a nonempty feasible set, there exists a unique global solution to maximize output each period. This solution, the indirect production function, can be written as a function of time t parameters:

$$f_t(N_t, Q_t) = N_t^{1-\alpha} Q_t^{1-\alpha} \quad (58)$$

where Q_t is an aggregated capital stock which is a weighted sum of the stocks of each vintage (Appendix 1). The clay-clay and putty-clay models maximize the output of the same Cobb-Douglas function, but subject to a fixed capital-labor ratio, r_v . Therefore, output corresponding to each vintage each period is:

$$f_{tv}(N_{tv}, K_{tv}) = \min(a_{tv} N_{tv}, d_{tv} K_{tv}) \quad (59)$$

where $a_{tv} = d_t A_v r_v^\alpha$ and $d_{tv} = d_t A_v r_v^{\alpha-1}$ measure the productivity of labor and capital, respectively.² In the clay-clay model, a_{tv} and d_{tv} are parameters, whereas in the putty-

² d_t and d_{tv} represent distinct concepts. d_t is disembodied technology corresponding to time period t whereas d_{tv} is the productivity of capital and includes both embodied (A_v) and disembodied technology.

clay model the central planner can choose r_v thereby affecting the levels of a_{tv} and d_{tv} .

Unlike the putty-putty model, there are no closed-form solutions for the indirect production functions in the clay-clay and putty-clay models.

The putty-putty and clay-clay models can be written as convex programming problems with objective functions that are strictly convex in consumption. Therefore there is a unique global solution for consumption. Consumption determines the time path of output in these two models and therefore there is also a unique global solution to the time path of output. The putty-clay model, however, cannot be written as a convex programming problem. While the objective functions in all three models are the same, the introduction of the capital-labor ratio choice variables in the putty-clay model leads to a non-convex feasible set. Therefore the same claim as to globalness and uniqueness of solutions cannot be made for the putty-clay model.

5. Initialization of Algorithms

Initialization of the models in Berger (2001) is simple. For the putty-clay model, first choose arbitrary capital-labor ratios for each vintage. Consequently for all three models, given a level of capital, labor, and parameter values in any given period, one can determine the maximum level of output. For the clay-clay and putty-clay models, maximizing output involves an algorithm which first allocates labor to the most productive capital until the capital-labor ratio is reached, and then allocating labor to the second most productive capital, continuing down the capital spectrum until either labor or capital run out. For the initial period, output can be set to a fraction of maximum output,

and consumption to a fraction of the chosen output. Output and consumption determine the following period's capital and therefore the procedure can be repeated for all of the periods. The slack variables should be set such that the optimality conditions are nearly satisfied given the initialization of the other primal variables. There is often no good intuition for the initial guess of the dual variables except for which will be zero (set these positive, but close to zero) and the relative scale of the others.

6. Putty-Putty Model

The putty-putty model examined in Berger (2001) is a discrete time version of the Phelps (1962) model with a finite time horizon. The Phelps model uses a constant rate of savings, and is divorced from dynamic optimization unlike this model. Technological growth is exogenous. The following model already incorporates the static production maximization, so that the production function given is the indirect production function (Appendix). The formal optimization problem is:

$$\begin{aligned}
 & \text{Minimize} && -\sum_{t=1}^T \beta^{t-1} \frac{1}{1-\gamma} (C_t^{1-\gamma} - 1) \\
 & \text{over} && C_t, Y_t, Q_t \quad \text{for } t=1 \dots T \\
 & \text{subject to:} && \\
 & && -C_t \leq 0 \\
 & && Y_t - d_t N_t^{1-\alpha} Q_t^\alpha \leq 0 \\
 & && C_t - Y_t \leq 0 \\
 & && Q_t - \bar{Q} \leq 0 \\
 & && Q_{t+1} - Q_t + A_t^{1/\alpha} (Y_t - C_t) \leq 0 \quad \text{for } t=1 \dots T-1
 \end{aligned} \tag{P_{pp}}$$

C_t is the chosen level of consumption in period t . Y_t is the chosen level of output in period t . Q_t is the level of the aggregate capital stock at time t . N_t is the given labor

stock each period. \bar{Q} is the given level of aggregate capital in the initial period. d_t is the given level of disembodied technology at time t . A_t is the given level of embodied technology corresponding to the vintage produced at time t . T is the given total number of time periods.

(P_{pp}) is a convex programming problem. It can be shown that (P_{pp}) is equivalent to a problem written in standard economic form, a maximization with the last four constraints written as equalities. The inequality constraints can be written as equalities because there are no transaction costs associated with the use of capital for production, or in carrying over capital and investment from previous periods into subsequent periods. Therefore, at a solution, the production and dynamic constraints will hold with equality.

Results for the putty-putty model were obtained for $T = 200$ so that the barrier form of the problem, as actually solved, had 1,800 primal variables (including 1,200 slack variables) and 1,200 dual variables (the number of constraints, including 600 non-negativity constraints). The value of the constant structure of the Jacobian of the optimality conditions to the equilibrium barrier formulation (22) is apparent in that the Newton direction can be found by inverting a 600 by 600 matrix, rather than the entire 3,000 by 3,000 Jacobian.

7. Clay-Clay Model

The clay-clay model, also known as the Leontief or fixed factor production model, is a discrete time variation of Solow, Tobin, Weizsäcker, Yaari (1966) with a finite time horizon. Unlike in the putty-putty model, there is no aggregate capital stock

because the embedded production maximization each period does not have a closed form solution. Thus, each vintage must be tracked separately. The capital stock in period t is a vector of $V+t-1$ elements, where V is the number of vintages available in the initial period ($t = 1$). The savings of period t becomes the quantity of capital of vintage $V+t$.

The production function, $Y_{tv} = \min(d_{tv}K_{tv}, a_{tv}N_{tv})$, where N_{tv} is labor in period t assigned to capital of vintage v , can be equivalently written as two inequality constraints.

$$Y_{tv} \leq a_{tv}N_{tv} \quad (60)$$

$$Y_{tv} \leq d_{tv}K_{tv} \quad (61)$$

The equivalency of the “min” function to the two inequality constraints arises because production will be maximized each period, for each vintage, given levels of labor and capital assigned to each vintage, and therefore (60) and/or (61) must hold with equality at a solution.

What does it mean if (60) does not hold strictly for time period t and vintage v , at a solution? It means that there is unproductive labor applied to vintage v . This can only be the case at a solution if there are no other vintages in that period to which the labor could be applied and be productive. If this is the case, then N_t , the labor available in period t , could be increased arbitrarily without changing the optimal time path of consumption and hence the optimal value of the optimization problem. Since the labor constraint each period, $\sum_{v=1}^{V+t-1} N_{tv} \leq N_t$, is only relevant when (60) is binding, we can substitute $\frac{Y_{tv}}{a_{tv}}$ for N_{tv} in the labor constraint, eliminating N_{tv} as a choice variable. This

yields the clay-clay model (a convex programming problem) solved in Berger (2001).

$$\begin{aligned}
& \text{Minimize} && -\sum_{t=1}^T \beta^{t-1} \frac{1}{1-\gamma} (C_t^{1-\gamma} - 1) \\
& \text{over} && C_t, Y_{tv}, K_{tv} \quad \text{for } t=1 \dots T \quad \text{for } v=1 \dots V+t-1 \\
& \text{subject to:} && \\
& && C_t \geq 0 \\
& && Y_{tv} \geq 0 \\
& && Y_{tv} - d_{tv} K_{tv} \leq 0 \\
& && \sum_{v=1}^{V+t-1} \frac{Y_{tv}}{a_{tv}} - N_t \leq 0 \\
& && C_t - \sum_{v=1}^{V+t-1} Y_{tv} \leq 0 \\
& && K_1 = \bar{K} \\
& && K_{t+1} = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} K_t + \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} C_t \\ Y_t \end{bmatrix} \quad \text{for } t=1 \dots T-1
\end{aligned} \tag{P_{cc}}$$

In Berger (2001), results were obtained for $T = 45$ and with no depreciation of capital. Depreciation can be easily incorporated into the model by changing the diagonal ones in the capital accumulation equation to fractional values. Consumption and output converge quickly to their equilibrium growth path in the clay-clay model, so 45 periods is ample time to observe the behavior of the system. Tracking each vintage each period greatly increases the number of variables in the model. The number of variables can be decreased somewhat by writing the problem in reduced form, substituting for K with a function of C , Y , and \bar{K} . The reduced form, equilibrium barrier formulation, with $T=45$ and $V = 2$, has 3,420 primal variables (including 2,295 slack variables), and 2,295 dual variables.

8. Putty-Clay

The putty-clay model is similar to the clay-clay model. The primary difference is that r_v , the capital-labor ratio for each vintage v , is a choice variable.

$$\begin{aligned}
 & \text{Maximize} \quad \sum_{t=1}^T \beta^{t-1} \frac{1}{1-\gamma} (C_t^{1-\gamma} - 1) \\
 & \text{over} \quad C_t, Y_{tv}, K_{tv}, r_v \quad \text{for } t=1 \dots T \quad \text{for } v=1 \dots V+t-1 \\
 & \text{subject to:} \\
 & \quad C_t \geq 0 \\
 & \quad Y_{tv} \geq 0 \\
 & \quad Y_{tv} - \frac{A_{tv}}{r_v^{1-\alpha}} K_{tv} \leq 0 \\
 & \quad \sum_{v=1}^{V+t-1} \frac{Y_{tv}}{A_{tv} r_v^\alpha} - N_t \leq 0 \tag{P_{pc}} \\
 & \quad C_t - \sum_{v=1}^{V+t-1} Y_{tv} \leq 0 \\
 & \quad K_1 = \bar{K} \\
 & \quad K_{t+1} = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ 0 & \dots & 0 \end{bmatrix} K_t + \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} C_t \\ Y_t \end{bmatrix} \quad \text{for } t=1 \dots T-1
 \end{aligned}$$

Unlike the putty-putty and clay-clay models, the putty-clay model can not be written as a convex programming problem (the feasible set of is not a convex set). Therefore, there can be multiple solutions to the model, and they need not be global. Berger (2001) does in fact find multiple solutions to the putty-clay problem.

Finding the results for the putty-clay model is more difficult than for the previous models. Because the problem is not a convex programming problem, there is no guarantee that the non-interior path-following algorithm will converge. The algorithm is highly sensitive to the initial guess, and often gets mired away from a solution. In order

to find solutions, it was necessary to use a bootstrapping method in which a solution was found for a low value of T , and this solution was used as the basis for an initial guess to a slightly longer problem. This procedure was repeated until a solution to the desired length problem was found. Also, because r is to a fractional power, it must be restricted to positive values in the algorithm. However, limiting the step size to keep the capital-labor ratio positive caused the program to stop progressing as the step size rapidly approached zero. So the primal variable r_v was replaced with s_v , where $r_v = \bar{r}e^{-s_v}$, and \bar{r} is a user defined algorithm parameter.

Two types of solutions to the model were found. One type has r_v constant from an early time period through the terminal period. The results suggest that there may be a local minimum for every constant level of r_v up to some threshold level. To obtain solutions of this type, it is possible to obtain a solution to a shorter horizon model, e.g. $T=10$, find the constant r_v , and then treat the problem as a clay-clay model using the obtained capital-labor ratios. When a solution to the 45 period clay-clay model is found, it can be used as an initial guess for the putty-clay model (re-running the non-interior point method is necessary in order to get the correct early period values for the 45 period putty-clay model). The non-interior method then only takes a few iterations to arrive at a solution.

When \bar{r} is set sufficiently high, another solution type is obtained. This solution has r_v increasing over time. This solution type also required bootstrapping to attain solutions to longer time period problems. Unlike for the constant capital-labor ratio solution however, there is no way to use the more stable clay-clay model to help obtain

the full 45 period solution. The putty-clay model has more variables than the clay-clay model because of the addition of 46 $(V + T - 1)$ capital-labor ratio variables.

9. Conclusion

In Berger (2001) numerical results were obtained for the three basic utility-maximizing growth models with vintage capital. The fact that complete time paths of the key economic variables were obtained for these models demonstrates the potential of the algorithms discussed in this paper. Given the specified parameter values, the results for the putty-putty and clay-clay model are definitive since the models can be written as convex programming problems with strictly convex objectives, and therefore have unique global solutions. The putty-clay model, however, cannot be written as a convex programming problem and multiple local solutions were found. Numerical results had not been previously obtained for optimization, discrete time versions of these models.

It was also shown that for convex programming problems the Newton method will be well defined for both the interior-point and non-interior point path-following algorithms. The results for the putty-clay model demonstrate both the difficulty, but also the possibility of attaining results for general non-linear models. The non-interior path following algorithm was unstable when applied to this model, but using bootstrapping techniques, solutions were found for the model. The putty-clay model is a large scale general non-linear model with several complications including fractional exponents. For smaller scale or convex programming problems, the non-interior method may be equally or more effective than the interior point method.

Models in economics are becoming increasingly complex. It will not be possible to find closed form theoretical solutions to many of these models. Analyzing models which more accurately reflect the world in which we live will require the use of algorithms such as those outlined in this paper.

Appendix 1: Discrete-Time, Vintage Capital Cobb-Douglas Production

Maximization

Each instant, a static optimization takes place in which given the stocks of technology, labor, and capital, labor must be allocated to the different vintages of capital in order to maximize production. The formal problem is as follows:

$$\begin{aligned}
 &\text{Maximize } Y = \sum_{v=1}^V A_v N_v^{1-\alpha} K_v^\alpha && 0 < \alpha < 1 \\
 &\text{over } N_v \\
 &\text{subject to:} && (62) \\
 &N_v \geq 0 \\
 &\sum_{v=1}^V N_v \leq N
 \end{aligned}$$

V is the newest vintage available. A_v is the level of technology associated with vintage v . N_v , the sole choice variable, is the quantity of labor assigned to capital of vintage v . N is the total amount of labor available. K_v is the amount of capital of vintage v available.

It is obvious that a solution will have $N_v > 0$ because the marginal product of labor of each vintage goes to infinity as N_v goes to zero. Also, the total labor constraint will hold with equality since the objective function rewards the use of labor with each vintage and there is no cost of using labor. The lagrangian for this problem is:

$$L(N_v; \lambda) = \sum_{v=1}^V A_v N_v^{1-\alpha} K_v^\alpha + \lambda \left(N - \sum_{v=1}^V N_v \right) \quad (63)$$

A representative first order condition is:

$$\frac{\partial L}{\partial N_v} = (1-\alpha) A_v N_v^{-\alpha} K_v^\alpha - \lambda = 0 \quad (64)$$

An equality can be used instead of an inequality, because $N_v > 0$ at a solution. (64)

states that the marginal product of labor of each vintage must be equal to the shadow value of labor. Labor, being homogeneous, has a unique shadow value regardless of the vintage to which it is applied. Setting two first order conditions for different vintages equal to each other we obtain:

$$A_i N_i^{\alpha-1} K_i^\alpha = A_j N_j^{\alpha-1} K_j^\alpha \quad (65)$$

Solving for N_i

$$N_i = \left(\frac{A_i}{A_j} \right)^{\frac{1}{\alpha}} \left(\frac{K_i}{K_j} \right) N_j \quad (66)$$

This equation must hold true for every i given a fixed A_j , N_j , and K_j . Therefore,

summing over the N_i :

$$N = \sum_{i=1}^v N_i = \frac{N_j}{A_j^{\frac{1}{\alpha}} K_j} \sum_{i=1}^v A_i^{\frac{1}{\alpha}} K_i \quad (67)$$

Solving for N_j in terms of the given parameters we obtain a closed form solution for each choice variable:

$$N_j^* = \frac{N A_j^{\frac{1}{\alpha}} K_j}{\sum_{i=1}^v A_i^{\frac{1}{\alpha}} K_i} \quad (68)$$

Let:

$$Q = \sum_{i=1}^v A_i^{\frac{1}{\alpha}} K_i \quad (69)$$

Q can be viewed as an aggregated capital stock with the weights of each vintage a function of the technology inherent in each vintage.

Substituting N_j^* into the direct production function, we obtain the indirect production function, which is a function only of given parameters.

$$\begin{aligned}
 Y^*(A, N, K, \sigma_A, \sigma_N, \sigma_K) &= \sum_{v=1}^V A_v K_v^\alpha \left(\frac{N A_v^{\frac{1}{\alpha}} K_v}{Q} \right)^{1-\alpha} \\
 &= N^{1-\alpha} Q^{\alpha-1} \sum_{v=1}^V A_v^{\frac{1}{\alpha}} K_v \\
 &= N^{1-\alpha} Q^\alpha
 \end{aligned} \tag{70}$$

References

- Berger, Brett D. (2001). "Convergence in Neoclassical Vintage Capital Growth Models." International Finance Discussion Papers, Board of Governors of the Federal Reserve System, 713, November 2001.
- Bertsekas, Dimitri. P. (1999). *Nonlinear Programming, Second Edition*. Athena Scientific.
- Burke, J., Xu, S. (2000). "A Non-Interior Predictor-Corrector Path Following Algorithm for the Monotone Linear Complementarity Problem." *Mathematical Programming* 87, 113-130.
- Greenwood, J., Z. Hercowitz, and P. Krussel. (1997). "Long-Run Implications of Investment-Specific Technological Change." *The American Economic Review* 87, 342-362.
- Phelps, Edmund S. (1962). "The New View of Investment: A Neoclassical Analysis." *Quarterly Journal of Economics* 76, 548-567.
- Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.
- Solow, Robert M. (1956). "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70, 65-94.
- Solow, Robert M. (1959). "Investment and Technical Progress." *Stanford Symposium on Mathematical Methods in the Social Sciences*, 89-104.
- Solow, R., J. Tobin, C. von Weizsäcker, and M. Yaari. (1966). "Neoclassical Growth with Fixed Factor Proportions." *Review of Economic Studies* 33, 79-155.
- Wright, Stephen J. (1997). *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics.