Chapter 7
Multivariable Unconstrained Optimization: Applications

7.1 Competitive Firm Input Choices: Cobb-Douglas Technology The profit function of a firm with Cobb-Douglas production function in a competitive product and inputs markets is given by $q = L^\alpha K^\beta$

$$max\ \pi = pL^\alpha K^\beta - wL - rK$$

L&K are decision variables.

First-Order Sufficient Condition:

$$\left.\begin{aligned}\frac{\partial \pi}{\partial L} &= \alpha p L^{*\alpha-1}K^{*\beta} - w = 0\\ \frac{\partial \pi}{\partial K} &= \beta p L^{*\alpha}K^{*\beta-1} - r = 0\end{aligned}\right\} \Rightarrow \begin{cases}\boxed{VMP_L = w}\\ VMP_K = r\end{cases}$$

where VMP_L and VMP_K are the *values of marginal product of labor and capital*, respectively.

Second-Order Sufficient Condition: The Hessian at the critical point (L^*, K^*) is negative definite.

$$\boxed{\mathbf{H}(L^*, K^*)} = \begin{bmatrix}\alpha(\alpha-1)pL^{*\alpha-2}K^{*\beta} & \alpha\beta pL^{*\alpha-1}K^{*\beta-1}\\ \alpha\beta pL^{*\alpha-1}K^{*\beta-1} & \beta(\beta-1)pL^{*\alpha}K^{*\beta-2}\end{bmatrix}$$

< 0

Test the negative definiteness:

$$|\mathbf{H}_1| = \alpha(\alpha-1)pL^{*\alpha-2}K^{*\beta} < 0, \text{ if } 0 < \alpha < 1$$

$|H| =$

$$\begin{aligned}|\mathbf{H}_2| = |\mathbf{H}| &= \alpha(\alpha-1)pL^{*\alpha-2}K^{*\beta}\beta(\beta-1)pL^{*\alpha}K^{*\beta-2}\\ &\quad - \left(\alpha\beta pL^{*\alpha-1}K^{*\beta-1}\right)^2\\ &= \alpha(\alpha-1)\beta(\beta-1)p^2L^{*2\alpha-2}K^{*2\beta-2}\\ &\quad - \alpha^2\beta^2p^2L^{*2\alpha-2}K^{*2\beta-2} > 0,\end{aligned}$$

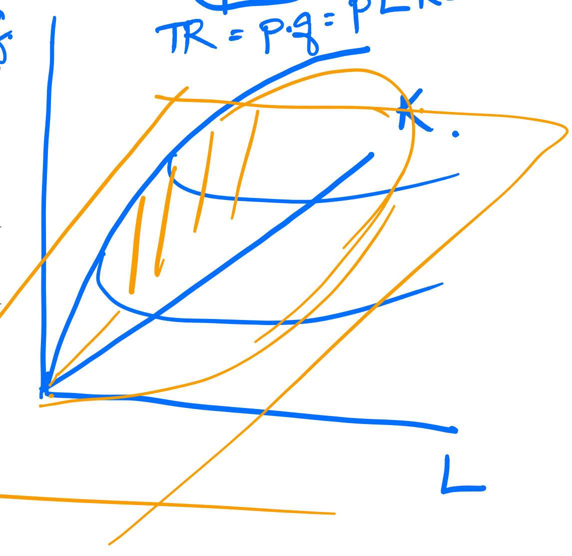
$$\Leftrightarrow (\alpha-1)(\beta-1) - \alpha\beta > 0, \text{ if } \beta > 0 \quad = 0$$

$$\Leftrightarrow \alpha\beta - \alpha - \beta + 1 - \alpha\beta = 1 - \alpha - \beta > 0.$$

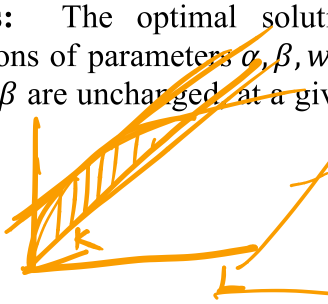
- That is, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$. The production function has to be homogeneous of degree less than 1. q
- That means the production function is decreasing returns to scale, and by Microeconomics, the marginal cost is positively sloped.

$q = L^\alpha K^\beta$ — decreasing returns to scale.
$\alpha + \beta < 1$

$p = 1$
$TR = p \cdot q = pL^\alpha K^\beta = L^\alpha K^\beta$



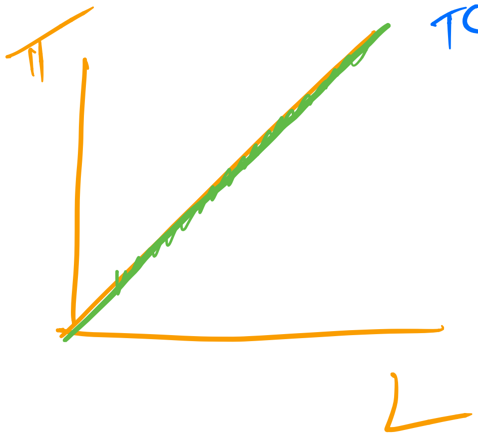
Comparative Static Analysis: The optimal solution (L^*, K^*) can be solved as functions of parameters α, β, w, r, and p. That is, assuming α and β are unchanged at a given

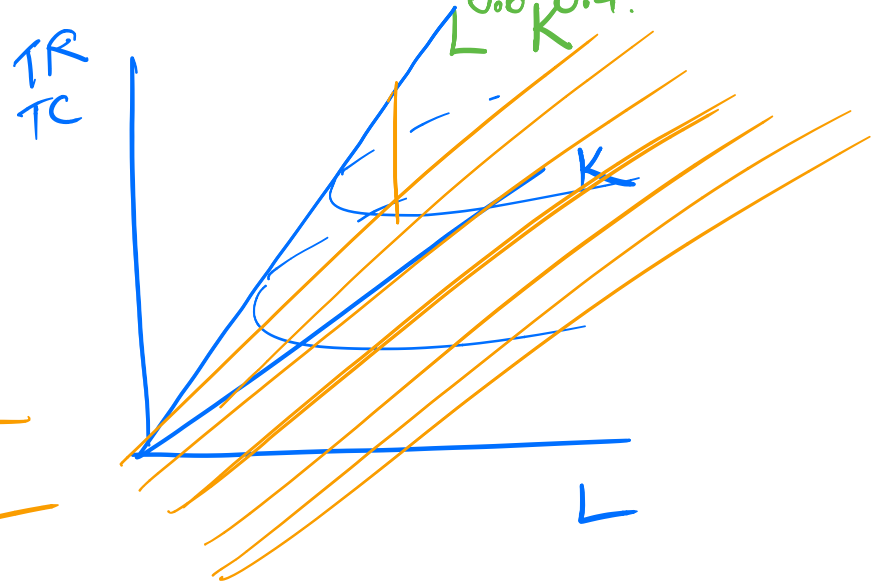


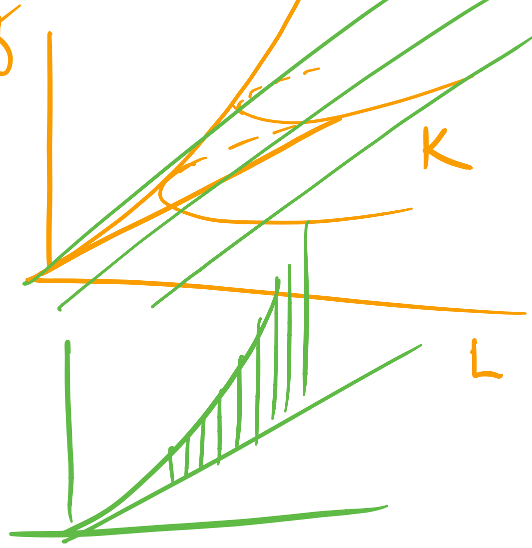
$\pi = TR - TC$
$p = 1 \quad = L^\alpha K^\beta - (wL - rK)$ linear

Constant Return to Scale $\alpha + \beta = 1$

$L^{0.6} K^{0.4}$.







particular (w_0, r_0, p_0) the first-order sufficient conditions can be written as an implicit functions of (L^*, K^*) as follows.

$$\mathbf{f}(L^*, K^*; w_0, r_0, p_0) = \begin{bmatrix} f^1(L^*, K^*; w_0, r_0, p_0) \\ f^2(L^*, K^*; w_0, r_0, p_0) \end{bmatrix} = \begin{bmatrix} \alpha p_0 L^{*\alpha_0 - 1} K^{*\beta_0} - w_0 \\ \beta p_0 L^{*\alpha_0} K^{*\beta_0 - 1} - r_0 \end{bmatrix} = \mathbf{0}.$$

The Implicit Function Theorem applies here because

$$\nabla_{\begin{bmatrix} L \\ K \end{bmatrix}} \mathbf{f}(L^*, K^*; w_0, r_0, p_0) = \mathbf{H}(L^*, K^*; w_0, r_0, p_0),$$

$|H| > 0$ – nonzero : H is nonsingular.

is nonsingular because the Hessian is negative definite. We have

a) there are functions $L(w, r, p)$ and $K(w, r, p)$ such that

$\nabla_{\begin{bmatrix} L \\ K \end{bmatrix}} \pi = 0$

$$\mathbf{f}(L(w, r, p), K(w, r, p); w, r, p) = \mathbf{0},$$

for $|w - w_0| < \varepsilon$, $|r - r_0| < \varepsilon$, and $|p - p_0| < \varepsilon$ for some $\varepsilon > 0$.

b) $L(w_0, r_0, p_0) = L^*$ and $K(w_0, r_0, p_0) = K^*$

c) The gradient

$$\nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \begin{bmatrix} L(w_0, r_0, p_0) \\ K(w_0, r_0, p_0) \end{bmatrix} = -\left[\nabla_{\begin{bmatrix} L \\ K \end{bmatrix}} \mathbf{f}(L^*, K^*; w_0, r_0, p_0)\right]^{-1} \nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \mathbf{f}(L^*, K^*; w_0, r_0, p_0)$$

2x3

$$= -\mathbf{H}(L^*, K^*; w_0, r_0, p_0)^{-1} \nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \mathbf{f}(L^*, K^*; w_0, r_0, p_0)$$

$$= -\begin{bmatrix} \alpha_0(\alpha_0 - 1) p_0 L^{*\alpha_0 - 2} K^{*\beta_0} & \alpha_0 \beta_0 p_0 L^{*\alpha_0 - 1} K^{*\beta_0 - 1} \\ \alpha_0 \beta_0 p_0 L^{*\alpha_0 - 1} K^{*\beta_0 - 1} & \beta_0(\beta_0 - 1) p_0 L^{*\alpha_0} K^{*\beta_0 - 2} \end{bmatrix}^{-1} \begin{bmatrix} -1 & 0 & \alpha_0 L^{*\alpha_0 - 1} K^{*\beta_0} \\ 0 & -1 & \beta_0 L^{*\alpha_0} K^{*\beta_0 - 1} \end{bmatrix}$$

2x2 2x3

2x3.

and by Cramer's Rule

$$\frac{\partial L(w_0, r_0, p_0)}{\partial w} = -\frac{\begin{vmatrix} -1 & \alpha_0 \beta_0 p_0 L^{*\alpha_0 - 1} K^{*\beta_0 - 1} \\ 0 & \beta_0(\beta_0 - 1) p_0 L^{*\alpha_0} K^{*\beta_0 - 2} \end{vmatrix}}{\begin{vmatrix} \alpha_0(\alpha_0 - 1) p_0 L^{*\alpha_0 - 2} K^{*\beta_0} & \alpha_0 \beta_0 p_0 L^{*\alpha_0 - 1} K^{*\beta_0 - 1} \\ \alpha_0 \beta_0 p_0 L^{*\alpha_0 - 1} K^{*\beta_0 - 1} & \beta_0(\beta_0 - 1) p_0 L^{*\alpha_0} K^{*\beta_0 - 2} \end{vmatrix}}$$

$$= \frac{\beta_0(\beta_0 - 1) p_0 L^{*\alpha_0} K^{*\beta_0 - 2}}{|\mathbf{H}|} < 0.$$

We can similarly find all other partial derivatives.

HW Baldani, p. 216, #8.2, 8.3, 8.4

7.2 Competitive Firm Input Choices: General Production Technology

The same firm as in 7.1 but now with a generic production function $f(L,K)$ will maximize the profit function

$$\max \pi = p\boxed{f(L,K)} - wL - rK.$$

First-Order Sufficient Condition:

$$\left.\begin{aligned}\frac{\partial \pi}{\partial L} = pf_L(L^*,K^*) - w = 0\\ \frac{\partial \pi}{\partial K} = pf_K(L^*,K^*) - r = 0\end{aligned}\right\} \Longrightarrow \begin{cases} VMP_L = w \\ VMP_K = r \end{cases}$$

Second-Order Sufficient Condition:

<0

$$\mathbf{H}(L^*,K^*) = \begin{bmatrix} pf_{LL} & pf_{LK} \\ pf_{KL} & pf_{KK} \end{bmatrix}$$

$$|\mathbf{H}_1| = pf_{LL} < 0$$
$$|\mathbf{H}_2| = |\mathbf{H}| = p^2(f_{LL}f_{KK} - f_{LK}^2) > 0.$$

By Implicit Function Theorem

$$\nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \begin{bmatrix} L(w_0,r_0,p_0) \\ K(w_0,r_0,p_0) \end{bmatrix} = -\left[\nabla_{\begin{bmatrix} L \\ K \end{bmatrix}} \mathbf{f}(L^*,K^*;w_0,r_0,p_0)\right]^{-1} \nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \mathbf{f}(L^*,K^*;w_0,r_0,p_0)$$
$$= -\mathbf{H}(L^*,K^*;w_0,r_0,p_0)^{-1} \nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \mathbf{f}(L^*,K^*;w_0,r_0,p_0)$$
$$= -\begin{bmatrix} p_0 f_{LL} & p_0 f_{LK} \\ p_0 f_{KL} & p_0 f_{KK} \end{bmatrix}^{-1} \begin{bmatrix} -1 & 0 & f_L \\ 0 & -1 & f_K \end{bmatrix}$$

and by Cramer's Rule

$$\frac{\partial L(w_0,r_0,p_0)}{\partial w} = -\frac{\begin{vmatrix} -1 & p_0 f_{LK} \\ 0 & p_0 f_{KK} \end{vmatrix}}{\begin{vmatrix} p_0 f_{LL} & p_0 f_{LK} \\ p_0 f_{KL} & p_0 f_{KK} \end{vmatrix}}$$
$$= \frac{p_0 f_{KK}}{|\mathbf{H}|} < 0.$$

MP_K is decreasing with k

$f_{KK} < 0$. $\frac{dMP_K}{dK} < 0$.

- The demand for labor has a negative slope with respect to wage rate if the capital exhibits diminishing returns-- $f_{KK} < 0$ means the slope of the marginal product is negative.

HW Determine the signs of $\frac{\partial L(w_0,r_0,p_0)}{\partial r}$ and $\frac{\partial L(w_0,r_0,p_0)}{\partial p}$.

HW Baldani, p. 216, #8.5.

7.3 Multi-plant Firm A firm with n plants. — 1 market.

$TC_i(q_i) = C_i(q_i)$, q_i = quantity produced at plant i.

q_i, $i = 1, \dots, n$ decision variables.

$$TR = R(Q) = P(Q)Q$$

$$Q = \sum_{i=1}^{n} q_i$$

$$\pi(\mathbf{q}) = R(Q) - \sum_{i=1}^{n} C_i(q_i) = P(Q)Q - \sum_{i=1}^{n} C_i(q_i)$$

$$\boxed{\pi_i(\mathbf{q}^*)} = P'(Q^*)Q^* + P(Q^*) - \boxed{C_i'(q_i^*)} = 0 \quad i = 1,2, \dots, n.$$

The last equality is the first-order sufficient condition and it implies that the marginal cost of each plant is equal to the marginal revenue, i.e.,

same for all plant i

$$\underbrace{C_1'(q_1^*)}_{MC_1} = \underbrace{C_2'(q_2^*)}_{MC_2} = \cdots = \underbrace{C_n'(q_n^*)}_{MC_n}$$
$$= P'(Q^*)Q^* + P(Q^*)$$
$$= R'(Q^*) = \boxed{MR(Q^*)} = 100.$$

$MC_1 = 102.$

$MC_2 = 95.$

The Hessian is given by <0 >0

$$\mathbf{H}(\mathbf{q}^*) = \begin{bmatrix} R'' - C_1'' & R'' & \cdots & R'' \\ R'' & R'' - C_2'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R'' - C_n'' \end{bmatrix}$$

$R'' = R''(q_1^*, q_2^*, \dots, q_n^*)$

$C_1'' = C_1''(q_1^*) > 0$

$\vdots$

$C_n'' = C_n''(q_n^*) > 0.$

where $R'' = P''(Q^*)Q^* + 2P'(Q^*)$. The Hessian is negative definite if,

i odd $(-1)^i = (-1)$

$$|\mathbf{H}_1| = R'' - C_1'' < 0$$
$$|\mathbf{H}_2| = (R'' - C_1'')(R'' - C_2'') - R''^2$$
$$= C_1''C_2'' - R''(C_1'' + C_2'') > 0$$
$$(-1)^i|\mathbf{H}_i| > 0, i = 3,4, \dots, n.$$

n

$C_i'' = \dfrac{d\, C_i'}{dq_i} = \dfrac{d\, MC_i}{dq_i} > 0.$ MC_i is increasing.

C

For perfect competition, so $\boxed{R'' = 0}$ and

$$\mathbf{H}(\mathbf{q}^*) = \begin{bmatrix} -C_1'' & 0 & \cdots & 0 \\ 0 & -C_2'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & -C_n'' \end{bmatrix}$$ — neg def.

Baldani.

which is negative definite if $C_i'' > 0, i = 1,2, \dots, n$.

For monopoly, this is not necessarily true if the plants have increasing return to scale where ($C_i'' < 0$).

HW Show that the Hessian \

$$\mathbf{H}(\mathbf{q}^*) = \begin{bmatrix} R'' - C_1'' & R'' & \cdots & R'' \\ R'' & R'' - C_2'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R'' - C_n'' \end{bmatrix}$$

neg def ?

is not necessarily negative definite if $C_i'' < 0$ for some $i = 1,2,\ldots,n$. (Hint: Show for $n = 2$)

small counterexample

HW Show that the Hessian $\mathbf{H}(\mathbf{q}^*)$ above is negative definite if $R_i'' < 0$ and $C_i'' > 0$, $i = 1,2,\ldots,n$.

Solution: If we assume instead that $R_i'' > 0$ and $C_i'' > 0$ then we can equivalently show that

$$\mathbf{H}(\mathbf{q}^*) = \begin{bmatrix} R'' + C_1'' & R'' & \cdots & R'' \\ R'' & R'' + C_2'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R'' + C_n'' \end{bmatrix}$$

$n \times n.$

is positive definite. We will prove by induction. If $n = 2$, then $|\mathbf{H}_1| = R'' + C_1'' > 0$ and

$$\begin{aligned} |\mathbf{H}_2| &= (R'' + C_1'')(R'' + C_2'') - R''^2 \\ &= C_1''C_2'' + R''(C_1'' + C_2'') > 0. \end{aligned}$$

Note that $|\mathbf{H}_2| > 0$ even when $C_1'' > 0$. We can then state the induction hypothesis that $|\mathbf{H}(\mathbf{q}^*)_{n\times n}| > 0$ when $R_i'' > 0, C_1'' > 0$ and $C_i'' > 0$, $i = 2,3,\ldots,n$. We need to prove

$$\left|\mathbf{H}(\mathbf{q}^*)_{(n+1)\times(n+1)}\right| = \begin{vmatrix} R'' + C_1'' & R'' & \cdots & R'' \\ R'' & R'' + C_2'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R'' + C_{n+1}'' \end{vmatrix} > 0.$$

By a property of determinant,

$$\left|\mathbf{H}(\mathbf{q}^*)_{(n+1)\times(n+1)}\right| = \begin{vmatrix} C_1'' & R'' & \cdots & \cdots & R'' \\ -C_2'' & R'' + C_2'' & \cdots & \cdots & \vdots \\ 0 & R'' & \ddots & \cdots & \vdots \\ \vdots & \cdots & \cdots & R'' + C_n'' & R'' \\ 0 & \cdots & \cdots & R'' & R'' + C_{n+1}'' \end{vmatrix}$$

$$= C_1'' \begin{vmatrix} R'' + C_2'' & R'' & \cdots & R'' \\ R'' & R'' + C_3'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R'' + C_{n+1}'' \end{vmatrix}$$

Proof by induction

To prove statement involving integer n.

$$+(-1)^{2+1}(-C_2'') \begin{vmatrix} R'' & R'' & \cdots & R'' \\ R'' & R'' + C_3'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R'' + C_{n+1}'' \end{vmatrix},$$

which is positive by the induction hypothesis.

- If we assume the functional forms of TR and TC explicitly with parameters, we can perform the sensitivity analysis using the Implicit Function Theorem.

HW Baldani, p. 217, #8.9

7.4 Multi-Market Monopoly A monopoly has two separate markets with similar demands $D_1(q_1) = \alpha P(q_1)$ and $D_2(q_2) = P(q_2)$, where q_1 and q_2 are quantities sold in the two markets. Assume that $\alpha > 1$ so that market 1 is more important. The total revenues earned are:

2 markets
1 seller (monopoly)

$$R_1(q_1) = \alpha P(q_1)q_1$$
$$R_2(q_2) = P(q_2)q_2.$$

The total cost of output $\boxed{Q = q_1 + q_2}$ is

$$\boxed{TC(Q) = C(Q) + tq_2,}$$

1 factory:

n factories
k markets

no. of decision variables =
n+k ?
n·k ?

where \$$t$ is the extra cost per unit to sell in market 2. The profit function is thus

$$\pi(q_1, q_2) = R_1(q_1) + R_2(q_2) - C(Q) - tq_2$$
$$= \alpha P(q_1)q_1 + P(q_2)q_2 - C(Q) - tq_2.$$

First-order Sufficient Condition:

$$\pi_1(q_1^*, q_2^*) = R_1'(q_1^*) - C'(q_1^* + q_2^*)$$
$$= \alpha P'(q_1^*)q_1^* + \alpha P(q_1^*) - C'(q_1^* + q_2^*) = 0$$
$$\pi_2(q_1^*, q_2^*) = R_2'(q_2^*) - C'(q_1^* + q_2^*) - t$$
$$= P'(q_2^*)q_2^* + P(q_2^*) - C'(q_1^* + q_2^*) - t = 0$$

plants	Market A	Market B
1	q_{1A}	q_{1B}
2	q_{2A}	q_{2B}
3	q_{3A}	q_{3B}

2×3 variables

We have $q_1^* > q_2^*$. (Why?)

Second-order Sufficient Condition: The Hessian is given by

$$\mathbf{H}(q_1^*, q_2^*) = \begin{bmatrix} R_1'' - C'' & -C'' \\ -C'' & R_2'' - C'' \end{bmatrix}$$

where

q_{ij} = amount to produce at plant i and send to market j.

$$R_1'' = \alpha P''(q_1^*)q_1^* + 2\alpha P'(q_1^*)$$
$$R_2'' = P''(q_2^*)q_2^* + 2P'(q_2^*).$$

Test of negative definiteness of the Hessian: If $R_1'' < 0$, $R_2'' < 0$ and $C'' > 0$.

q_1, q_2, q_3

q_A, q_B.

$$|\mathbf{H}_1(q_1^*, q_2^*)| = R_1'' - C'' < 0$$
$$|\mathbf{H}(q_1^*, q_2^*)| = (R_1'' - C'')(R_2'' - C'') - C''^2$$

$$= R_1'' R_2'' - R_1'' C'' - R_2'' C'' > 0.$$

Comparative Static Analysis: Write the first-order condition as implicit functions.

$$\nabla_{\mathbf{q}} \pi(q_1^*, q_2^*; \alpha_0, t_0) = \begin{bmatrix} \pi_1(q_1^*, q_2^*; \alpha_0, t_0) \\ \pi_2(q_1^*, q_2^*; \alpha_0, t_0) \end{bmatrix}$$
$$= \begin{bmatrix} \alpha_0 P'(q_1^*) q_1^* + \alpha_0 P(q_1^*) - C'(q_1^* + q_2^*) \\ P'(q_2^*) q_2^* + \alpha_0 P(q_2^*) - C'(q_1^* + q_2^*) - t_0 \end{bmatrix} = \mathbf{0}$$

Assuming $\nabla_{\mathbf{q}}^2 \pi(q_1^*, q_2^*; \alpha_0, t_0) = \mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)$ being a nonsingular matrix, the Implicit Function Theorem yields

$$\nabla_{\left[\begin{smallmatrix}\alpha \\ t\end{smallmatrix}\right]} \mathbf{q}(\alpha_0, t_0)$$
$$= -\left[\nabla_{\mathbf{q}}^2 \pi(q_1^*, q_2^*; \alpha_0, t_0)\right]^{-1} \nabla_{\left[\begin{smallmatrix}\alpha \\ t\end{smallmatrix}\right]} \left(\nabla_{\mathbf{q}} \pi(q_1^*, q_2^*; \alpha_0, t_0)\right)$$
$$= -\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)^{-1} \begin{bmatrix} P'(q_1^*) q_1^* + P(q_1^*) & 0 \\ 0 & -1 \end{bmatrix}$$
$$= -\begin{bmatrix} R_1'' - C'' & -C'' \\ -C'' & R_2'' - C'' \end{bmatrix}^{-1} \begin{bmatrix} R_1'/\alpha & 0 \\ 0 & -1 \end{bmatrix}$$

By Cramer's Rule, if $P'(q_1^*) q_1^* + P(q_1^*) > 0$ and $C'' > 0$

$$\frac{\partial q_1^*}{\partial \alpha} = -\frac{\begin{vmatrix} \frac{R_1'}{\alpha} & -C'' \\ 0 & R_2'' - C'' \end{vmatrix}}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|} = -\frac{\frac{R_1'}{\alpha}(R_2'' - C'')}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|} > 0,$$
$$\frac{\partial q_1^*}{\partial t} = -\frac{\begin{vmatrix} 0 & -C'' \\ -1 & R_2'' - C'' \end{vmatrix}}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|} = \frac{C''}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|} > 0$$

Thus,

$$dq_1^* = \frac{\partial q_1^*}{\partial \alpha} d\alpha + \frac{\partial q_1^*}{\partial t} dt$$
$$= \frac{-\frac{R_1'}{\alpha}(R_2'' - C'') d\alpha + C'' dt}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|}.$$

and similarly,

$$dq_2^* = \frac{\partial q_2^*}{\partial \alpha} d\alpha + \frac{\partial q_2^*}{\partial t} dt$$
$$= \frac{\frac{R_1'}{\alpha} C'' d\alpha + (R_1'' - C'') dt}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|}.$$

We have

$$dQ^* = dq_1^* + dq_2^* = \frac{-\frac{R_1'}{\alpha}R_2'' + R_1'' dt}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|}.$$

HW Baldani, p. 217, #8.10, 8.11

$\beta_0, \beta_1, \ldots, \beta_k$ are unknowns.

7.5 Statistical Estimation: Linear Regression

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \begin{matrix} = \beta_0 + \beta_1 x_1^1 + \beta_2 x_2^1 + \cdots + \beta_k x_k^1 + \varepsilon_1 \\ \vdots \\ = \beta_0 + \beta_1 x_1^n + \beta_2 x_2^n + \cdots + \beta_k x_k^n + \varepsilon_n \end{matrix}$$

$\mathbf{y}$ - vector of constant - data collected.

Recall matrix differentiation,

$\nabla A\mathbf{x} = A$

$$\nabla \mathbf{c}^{\mathrm{T}}\mathbf{x} = \mathbf{c}$$
$$\nabla \alpha \mathbf{x}^{\mathrm{T}}\mathbf{x} = 2\alpha\mathbf{x}$$
$$\nabla \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = 2\mathbf{A}\mathbf{x},$$

where $\mathbf{A}$ is a symmetric matrix.

$$\mathbb{X} = \begin{bmatrix} 1 & x_1^1 & \cdots & x_k^1 \\ 1 & \vdots & & \vdots \\ \vdots & & & \\ 1 & x_1^n & \cdots & x_k^n \end{bmatrix} \text{ - constant}, \quad n \times (k+1)$$

Linear regression model: *Least Squares* The dependent variable y is determined linearly by the independent variables $x_j, j = 1,2,\ldots,k,$ with some random error ε. Suppose there are n observations, we have for $i = 1,2,\ldots,n,$

dependent. independent

$$y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \cdots + \beta_k x_i^k + \varepsilon_i,$$

Consumption income size . . . .

$$\mathbb{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \in \mathbb{R}^{k+1} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

where $\beta_j, j = 0,1,2,\ldots,k,$ are unknown parameters. In matrix form,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

Want to estimate $\mathbb{B}$ by a vector $\mathbb{b}$ so we have

$$\hat{y} = \mathbb{X}\mathbb{b}$$

where $\mathbf{y}, \boldsymbol{\varepsilon} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^{k+1}$, and $\mathbf{X} \in \mathbb{R}^{n\times(k+1)}$. If we estimate $\boldsymbol{\beta}$ by some $\mathbf{b} \in \mathbb{R}^{k+1}$, the estimate of $\mathbf{y}$ is thus $\hat{\mathbf{y}} = \mathbf{Xb}$, and the error of estimation of $\mathbf{y}$ is $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$. The determination of $\mathbf{b}$ by the Least Squares Method is the choice of $\mathbf{b}$ such that the sum of squares of the error of estimation $SSR = \mathbf{e}^{\mathrm{T}}\mathbf{e} = \sum_{i=1}^{n} e_i^2$ is minimized. The minimization problem is thus given by

$$e = y - \hat{y} \text{ - error vector.}$$

$\mathbb{b}$ decision variables.

$$\min_{\mathbf{b}\in\mathbb{R}^{k+1}} f(\mathbf{b}) = \mathbf{e}^{\mathrm{T}}\mathbf{e} = (\mathbf{y} - \hat{\mathbf{y}})^{\mathrm{T}}(\mathbf{y} - \hat{\mathbf{y}})$$
$$= (\mathbf{y} - \mathbf{Xb})^{\mathrm{T}}(\mathbf{y} - \mathbf{Xb})$$
$$= \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\mathbf{Xb} + \mathbf{b}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Xb}.$$

$SSR = e^T e$

By the first-order sufficient condition, the critical solution is the solution that

$$\nabla f(\mathbf{b}) = -2\mathbf{X}^{\mathrm{T}}\mathbf{y} + 2\mathbf{X}^{\mathrm{T}}\mathbf{Xb} = \mathbf{0}$$

$$\nabla \underset{1\times n \;\; n\times k+1}{y^T \mathbb{X}}\, b = \left[y^T \mathbb{X}\right]^T_{1\times k+1} = \mathbb{X}^T y$$

and thus the critical solution is given by

$$\hat{\mathbf{b}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y},$$

$\hat{b}$ exists and unique.

$$\underset{(k+1)\times n}{\mathbb{X}^T}\ \underset{n\times(k+1)}{\mathbb{X}} \quad (k+1)\times k+1$$

$$\nabla b^T(\mathbb{X}^T\mathbb{X})b = 2(\mathbb{X}^T\mathbb{X})b$$

$$(\mathbb{X}^T\mathbb{X})^T = \mathbb{X}^T\mathbb{X}$$

$-\cancel{2}X^T y + \cancel{2}X^T X \hat{b} = 0$

$\boxed{X^T X \hat{b} = X^T y}$ critical point.

$\hat{b} = (X^T X)^{-1} X^T y$

if $X^T X$ is nonsingular

and the second-order sufficient condition requires that

$$\nabla^2 f(\hat{\mathbf{b}}) = \boxed{2\mathbf{X}^T\mathbf{X},} \text{ pos def.}$$

be positive definite. The square symmetric matrix $\mathbf{X}^T\mathbf{X}$ is always positive semidefinite. Why? Now, if it is also positive definite, the solution $\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, as given by the first-order sufficient condition is uniquely defined because $(\mathbf{X}^T\mathbf{X})^{-1}$ exists (Why?), is a strict local minimum point. Can we say that is a strict global minimum?

$X^T X$ - always pos semidef

HW What will happen if one of the independent variable is just a linear combination of some of the other independent variable? For example, what if $x^1 = 1.5x^2$?

proof $X^T X$ always pos semidef.

by Def $X^T X$ is pos semi def if

$$\underbrace{d^T(X^T X)d}\geq 0 \quad \text{for any } d \in \mathbb{R}^{k+1}$$

Let $z = Xd$

$$d^T X^T X d = z^T z = \sum_{i=1}^{k+1} z_i^2 \geq 0.$$

want $X^T X$ to be pos def.

$$d^T X^T X d > 0 \quad \text{for any } d \in \mathbb{R}^{k+1},\ d \neq 0$$

$X^T X$ is pos def under a very simple condition!

col of X are linearly indep.

$d^T X^T X d = z^T z = 0$ only when. $z = 0$.

only when $Xd = 0$

$$X = \begin{bmatrix} 1 & x_1^1 & \cdots & x_1^k \\ 1 & x_2^1 & & x_2^k \\ \vdots & \vdots & & \vdots \\ 1 & x_n^1 & & x_n^k \end{bmatrix} = \begin{bmatrix} \mathbb{1} & x^1 & x^2 & \cdots & x^k \end{bmatrix}$$

$$\mathbb{X}\mathbf{d} = \begin{bmatrix} \mathbb{1} & \mathbb{x}^1 & \mathbb{x}^2 & \cdots & \mathbb{x}^k \end{bmatrix} \begin{bmatrix} d_0 \\ d_1 \\ \vdots \\ d_k \end{bmatrix}$$

If we can find some $\mathbf{d}$ such that

$$= \underline{d_0 \mathbb{1} + d_1 \mathbb{x}^1 + d_2 \mathbb{x}^2 + \cdots + d_k \mathbb{x}^k} = \mathbb{0}$$

we can find a <u>linear combination</u> of columns of $\mathbb{X}$ to be equal to $\mathbb{0}$

$$\mathbb{X} = \begin{array}{cccc} d_0 & d_1 & d_2 & \\ \end{array}\left[\begin{array}{cccc} 1 & 20 & 2 & \cdots \\ 1 & 30 & 3 & \\ \vdots & 60 & 6 & \cdots \\ 1 & \vdots & & \cdots \end{array}\right] \qquad \mathbb{X} \left[\begin{array}{cccc} 1 & 20 & 5 & 25 \\ 1 & 30 & 7 & 37 \\ \vdots & 60 & 3 & 63 \\ \vdots & & & \\ 1 & & & \end{array}\right]$$

$d_0 = 0$, $d_3 = d_4 = \cdots = d_k = 0.$

$d_1 = 1$ $d_2 = -10$

If there exists $\mathbf{d} \neq \mathbb{0}$ such that $\mathbb{X}\mathbf{d} = \mathbb{0}$, we say <u>columns of $\mathbb{X}$ are linearly dependent.</u>

If there is no $\mathbf{d} \neq \mathbb{0}$ that will make $\mathbb{X}\mathbf{d} = \mathbb{0}$ we say <u>columns of $\mathbb{X}$ are linearly independent</u>

$$\underbrace{\mathbf{d}^T\mathbb{X}^T}_{\neq \mathbb{0}^T}\underbrace{\mathbb{X}\mathbf{d}}_{\neq \mathbb{0}} > 0 \text{ — pos def.}$$

If $\mathbb{X}^T\mathbb{X}$ is <u>pos</u> def, then $\mathbb{X}^T\mathbb{X}$ is nonsingular

(neg)

Since $\nabla^2 f(\mathbb{b}) = 2\mathbb{X}^T\mathbb{X}$ – does not involve $\mathbb{b}$.

$\nabla^2 f(\mathbb{b})$ is pos def for any $\mathbb{b}$

$\therefore$ $f(\mathbb{b})$ is strictly convex.

$\Rightarrow$ $\hat{\mathbb{b}}$ is strictly global min.