Cost Minimization and the Cost Function

Juan Manuel Puerta

October 5, 2009
So far we focused on profit maximization, we could look at a different problem, that is the cost minimization problem. This is useful for some reasons:

- Different look of the supply behavior of competitive firms
- But also, this way we can model supply behavior of firms that don’t face competitive output prices
- (Pedagogic) We get to use the tools of constrained optimization

**Cost Minimization Problem:** \( \min_x w x \) such that \( f(x) = y \)

Begin by setting-up the Lagrangian: \( \mathcal{L}(\lambda, x) = wx - \lambda (f(x) - y) \)

Differentiating with respect to \( x_i \) and \( \lambda \) you get the first order conditions,

\[
    w_i - \lambda \frac{\partial f(x^*)}{\partial x_i} = 0 \quad \text{for} \ i = 1, 2, \ldots, n
\]

\[
    f(x^*) = y
\]
Letting $\mathbf{D}f(\mathbf{x})$ denote the gradient of $f(\mathbf{x})$, we can write the $n$ derivative conditions in matrix notation as,

$$\mathbf{w} = \lambda \mathbf{D}f(\mathbf{x}^*)$$

Dividing the $i^{th}$ condition by the $j^{th}$ condition we can get the familiar first order condition,

$$\frac{w_i}{w_j} = \frac{\partial f(\mathbf{x}^*)}{\partial x_i} \frac{\partial f(\mathbf{x}^*)}{\partial x_j}, \text{ for } i, j = 1, 2, \ldots, n$$

(1)

1. This is the standard “isocost=slope of the isoquant” condition †

2. Economic intuition: What would happen if (1) is not an equality? †
Source: Varian, Microeconomic Analysis, Chapter 4, p. 51.
Second Order Conditions

- In our discussion above, we assume we “approach” the isoquant from below. Do you see that if your isoquant is such that the isocost approaches it from above there is a problem?

- Other way of saying this is: if we move along the isocost, we cannot increase output? Indeed, output should remain constant or be reduced.

- Assume differentiability and take a second-order Taylor approximation of $f(x_1 + h_1, x_2 + h_2)$ where $h_i$ are small changes in the input factors. Then,

$$f(x_1 + h_1, x_2 + h_2) \approx f(x_1, x_2) + \frac{\partial f(x_1, x_2)}{\partial x_1} h_1 + \frac{\partial f(x_1, x_2)}{\partial x_2} h_2 + (1/2)[\frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} h_1^2 + 2 \frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2} h_1 h_2 + \frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} h_2^2]$$
Second Order Conditions

- Since we assumed a move along the isocost,
  \[ w_1 h_1 + w_2 h_2 = 0 = \lambda (f_1 h_1 + f_2 h_2) \] where the last equality follows from using FOC \((w_i = \lambda f_i)\)

- But in order to be at an optimum,
  \[ f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) \leq 0, \] which means that
  \[
  \begin{pmatrix}
  h_1 & h_2
  \end{pmatrix}
  \begin{pmatrix}
  f_{11} & f_{12} \\
  f_{21} & f_{22}
  \end{pmatrix}
  \begin{pmatrix}
  h_1 \\
  h_2
  \end{pmatrix}
  \leq 0
  \] (2)
  for \(f_1 h_1 + f_2 h_2 = 0\)

- Generalizing to the n-factor case,
  \[ h'D^2f(x)h \leq 0 \] for all \(h\) satisfying \(wh = 0\)
  Where \(h = (h_1, h_2, \ldots, h_n)\) is a quantity vector (a column vector according to our convention) and \(D^2f(x)\) is the Hessian of the production function.

- Intuitively, FOC imply that the isocost is tangent to the isoquant. SOC imply that a move along the isocost results on a reduction.
• The second order conditions can also be expressed in terms of the Hessian of the Lagrangian

\[ D^2 \mathcal{L}(\lambda^*, x_1^*, x_2^*) = \begin{pmatrix}
\frac{\partial^2 \mathcal{L}}{\partial \lambda^2} & \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial x_1} & \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial x_2} \\
\frac{\partial^2 \mathcal{L}}{\partial x_1 \partial \lambda} & \frac{\partial^2 \mathcal{L}}{\partial x_1^2} & \frac{\partial^2 \mathcal{L}}{\partial x_1 \partial x_2} \\
\frac{\partial^2 \mathcal{L}}{\partial x_2 \partial \lambda} & \frac{\partial^2 \mathcal{L}}{\partial x_2 \partial x_1} & \frac{\partial^2 \mathcal{L}}{\partial x_2^2}
\end{pmatrix} \]

• Computing these derivatives for the case of our Lagrangian function \( \mathcal{L}(\lambda, x) = wx - \lambda(f(x) - y) \)
The resulting is the **Bordered Hessian**

\[
\mathbf{D}^2 \mathcal{L}(\lambda^*, x_1^*, x_2^*) = \begin{pmatrix}
0 & -f_1 & -f_2 \\
-f_1 & -\lambda f_{11} & -\lambda f_{12} \\
-f_2 & -\lambda f_{21} & -\lambda f_{22}
\end{pmatrix}
\]

It turns out that the sufficient conditions stated in (2) are satisfied with strict inequality *if and only if* the determinant of the bordered hessian is **negative**. Similarly, if you have \( n \) factors, the bordered Hessians for the \( n \)-cases should be negative.
Difficulties

- For each choice of \( w \) and \( y \), there should be an optimum \( x^* \) that minimizes the cost of producing \( y \). This is the **Conditional Factor Demand** (Cf. factor demands in profit maximization).

- Similarly, the **Cost Function** is the function that gives the minimum cost of producing \( y \) at the factor prices \( w \).

\[
c(w, y) = wx(w, y)
\]
As in the profit maximization case, there could be cases in which the first order conditions would not work

1. Technology not representable by a differential production function (e.g. Leontieff)
2. We are assuming interior solution, i.e. that all the inputs are used in a strictly positive amount. Otherwise, we have to modify the conditions according to Kuhn-Tucker,
   \[ \lambda \frac{\partial f(x^*)}{\partial x_i} - w_i \leq 0 \text{ with strict equality if } x_i > 0 \]
3. The third issue concerns the existence of the optimizing bundle. The cost function, unlike the profit function, will always achieve a minimum. This follows from the fact that a continuous function achieves a minimum and a maximum on a compact (close and bounded) set. (more on that on the next slide)
4. The fourth problem is the issue of uniqueness. As we saw, calculus often ensures that a local maximum is achieved. For finding global maxima you have to make extra assumptions, namely \( V(y) \) convex.
More on why existence of a solution would not be a problem in the cost minimization case?

Because we are minimizing a continuous function on a close and bounded set. To see this, $wx$ is certainly continuous and $V(y)$ is closed by assumption (regularity assumption). Boundedness could be proved easily. Assume an arbitrary $x'$, then the minimal cost bundle must have a lower cost, $wx \leq wx'$. But then we can restrict to a subset \{x in V(y):wx \leq wx'\}, which is bounded so long $w \gg 0$. 
Some Examples of Cost Minimization

- Cobb Douglas with 2 inputs, $f(x_1, x_2) = A x_1^a x_2^b$. †
- CES, $f(x_1, x_2) = (x_1^\rho + x_2^\rho)^{1/\rho}$. Homework!
- Leontieff, $f(x_1, x_2) = \min\{ax_1, bx_2\}$. †
- Linear, $f(x_1, x_2) = ax_1 + bx_2$. Illustration of the Kuhn-Tucker conditions. †
2-input case

- In the usual fashion, the conditional factor demand function imply the following identities,

\[ f(x(w,y)) \equiv y \]

\[ w - \lambda Df(x(w,y)) \equiv 0 \]

- For the simpler 1-output, 2-input case, FOC imply

\[ f(x_1(w_1,w_2,y), x_2(w_1,w_2,y)) \equiv y \]

\[ w_1 - \lambda \frac{\partial f(x_1(w_1,w_2,y), x_2(w_1,w_2,y))}{\partial x_1} \equiv 0 \]

\[ w_2 - \lambda \frac{\partial f(x_1(w_1,w_2,y), x_2(w_1,w_2,y))}{\partial x_2} \equiv 0 \]
As we did with the FOC of the profit maximization problem, we can differentiate these identities with respect to the parameters, e.g. $w_1 \dagger$

$$\frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial w_1} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial w_1} \equiv 0$$

$$1 - \lambda\left[\frac{\partial^2 f}{\partial x_1^2} \frac{\partial x_1}{\partial w_1} + \frac{\partial^2 f}{\partial x_1 \partial x_2} \frac{\partial x_2}{\partial w_1}\right] - \frac{\partial f}{\partial x_1} \frac{\partial \lambda}{\partial w_1} \equiv 0$$

$$0 - \lambda\left[\frac{\partial^2 f}{\partial x_2 \partial x_1} \frac{\partial x_1}{\partial w_1} + \frac{\partial^2 f}{\partial x_2^2} \frac{\partial x_2}{\partial w_1}\right] - \frac{\partial f}{\partial x_2} \frac{\partial \lambda}{\partial w_1} \equiv 0$$

Which can be written in matrix form as,

$$\left(\begin{array}{ccc} 0 & -f_1 & -f_2 \\ -f_1 & -\lambda f_{11} & -\lambda f_{21} \\ -f_2 & -\lambda f_{12} & -\lambda f_{22} \end{array}\right) \left(\begin{array}{c} \frac{\partial \lambda}{\partial w_1} \\ \frac{\partial \lambda}{\partial x_1} \\ \frac{\partial \lambda}{\partial x_2} \end{array}\right) \equiv \left(\begin{array}{c} 0 \\ -1 \\ 0 \end{array}\right)$$

Note that the matrix on the left is precisely the “Bordered Hessian”
Recall the Cramer’s Rule, we can use it to solve for $\partial x_i / \partial w_1$:

$$
\frac{\partial x_1}{\partial w_1} = \frac{1}{\det \begin{vmatrix}
0 & 0 & -f_2 \\
-f_1 & -1 & -\lambda f_{21} \\
-f_2 & 0 & -\lambda f_{22}
\end{vmatrix}} \det \begin{vmatrix}
0 & -f_1 & -f_2 \\
-f_1 & -\lambda f_{11} & -\lambda f_{21} \\
-f_2 & -\lambda f_{12} & -\lambda f_{22}
\end{vmatrix}
$$

Solving the determinant on the top, and letting $H$ denote the lower determinant,

$$
\frac{\partial x_1}{\partial w_1} = \frac{f_2^2}{H} < 0
$$

In order to satisfy SOC, $H < 0$, which means that the conditional factor demand has a negative slope.
Similarly, you can use Cramer’s rule to solve for \( \frac{\partial x_2}{\partial w_1} \),

\[
\frac{\partial x_2}{\partial w_1} = \frac{\begin{vmatrix} 0 & -f_1 & 0 \\ -f_1 & -\lambda f_{11} & -1 \\ -f_2 & -\lambda f_{12} & 0 \end{vmatrix}}{\begin{vmatrix} 0 & -f_1 & -f_2 \\ -f_1 & -\lambda f_{11} & -\lambda f_{21} \\ -f_2 & -\lambda f_{12} & -\lambda f_{22} \end{vmatrix}}
\]

Carrying out the calculations,

\[
\frac{\partial x_2}{\partial w_1} = \frac{-f_2 f_1}{H} > 0
\]

Similarly, you can differentiate the identities above with respect to \( w_2 \) to get

\[
\begin{pmatrix} 0 & -f_1 & -f_2 \\ -f_1 & -\lambda f_{11} & -\lambda f_{21} \\ -f_2 & -\lambda f_{12} & -\lambda f_{22} \end{pmatrix} \begin{pmatrix} \frac{\partial \lambda}{\partial w_2} \\ \frac{\partial x_1}{\partial w_2} \\ \frac{\partial x_2}{\partial w_2} \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}
\]
And using the Cramer’s rule again, you can obtain

\[
\frac{\partial x_2}{\partial w_1} = \frac{-f_1f_2}{H} > 0
\]

Compare the expressions for \(\frac{\partial x_1}{\partial w_2}\) and \(\frac{\partial x_2}{\partial w_1}\). You will notice that as in the case of the factor demand functions, there is a symmetry effect.

In the 2-factor case, \(\frac{\partial x_i}{\partial w_j} > 0\), means that factors are always substitutes.

Of course, this analysis is readily extended to the n-factor case. As in the case of the profit maximization problem, it is better to use matrix notation.
n-input case

- The first order conditions for cost minimization are \(^1\)
  \[
  f(x(w)) \equiv y
  
  w - \lambda Df(x(w)) \equiv 0
  
  \]

- Differentiating these identities with respect to \(w\),
  \[
  Df(x(w)) Dx(w) = 0
  
  I - \lambda D^2f(x(w)) Dx(w) - Df(x(w)) D\lambda(w) = 0
  
  \]

- Rearranging this expression,
  \[
  \begin{pmatrix}
  0 & -Df(x(w)) \\
  -Df(x(w))' & -\lambda D^2f(x(w))
  \end{pmatrix}
  \begin{pmatrix}
  D\lambda(w) \\
  Dx(w)
  \end{pmatrix}
  = - \begin{pmatrix}
  0 \\
  I
  \end{pmatrix}
  
  \]

\(^1\)We omitted \(y\) as an argument as it is fixed.
Assuming a regular optimum so that the Hessian is non-degenerate. Then pre-multiplying each side by the inverse of the Bordered Hessian we obtain,

\[
\begin{pmatrix}
D\lambda(w) \\
Dx(w)
\end{pmatrix} = \begin{pmatrix}
0 & Df(x(w)) \\
Df(x(w))' & \lambda D^2 f(x(w))
\end{pmatrix}^{-1} \begin{pmatrix} 0 \\
I
\end{pmatrix}
\]

From this expression, it follows that, since the hessian is symmetric, the cross-price effects are symmetric.

Also, It can be shown that the substitution matrix is negative semi-definite.
The cost function

The **cost function** tells us the minimum cost of producing a level of output given certain input prices.

The cost function can be expressed in terms of the conditional factor demands we talked about earlier

\[ c(w, y) \equiv wx(w, y) \]

**Properties of the cost function.** As with the profit function, there are a number of properties that follow from cost minimization. These are:

1. Nondecreasing in \( w \). If \( w' \geq w \), then \( c(w', y) \geq c(w, y) \)
2. Homogeneous of degree 1 in \( w \). \( c(tw, y) = tc(w, y) \) for \( y > 0 \).
3. Concave in \( w \). \( c(tw + (1-t)w') \geq tc(w) + (1-t)c(w') \) for \( t \in [0, 1] \)
4. Continuous in \( w \). \( c(w, y) \) is a continuous function of \( w \), for \( w \gg 0 \)
Proof †:

- Non-decreasing
- Homogeneous of Degree 1
- Concave
- Continuous

Intuition for concavity of the cost function †
Proof †:

- Non-decreasing
- Homogeneous of Degree 1
- Concave
- Continuous

Intuition for concavity of the cost function †
Proof †:

- Non-decreasing
- Homogeneous of Degree 1
- Concave
- Continuous

Intuition for concavity of the cost function †
Proof †:

- Non-decreasing
- Homogeneous of Degree 1
- Concave
- Continuous

Intuition for concavity of the cost function †
Proof †:

- Non-decreasing
- Homogeneous of Degree 1
- Concave
- Continuous

Intuition for concavity of the cost function †
Shepard’s Lemma

**Shepard’s Lemma:** Let $x_i(w, y)$ be the firm’s conditional factor demand for input $i$. Then if the cost function is differentiable at $(w, y)$, and $w_i > 0$ for $i = 1, 2, 3, \ldots, n$ then,

$$x_i(w, y) = \frac{\partial c(w, y)}{\partial w_i} \text{ for } i = 1, 2, \ldots, n$$

**Proof †**

In general there are 4 approaches to proof and understand Shepard’s Lemma

1. Differentiate de identity and use FOC (Problem set 2)
2. Use the Envelope theorem directly (see next section)
3. A geometric argument †
4. A economic argument. At the optimum $x$, a small change in factor prices has a direct and an indirect effect. The indirect effect occurs through re-optimization of $x$ but this is negligible in the optimum. So we are left with the direct effect alone, and this is just equal to $x$. 
Envelope theorem for constrained optimization

- Shepard’s lemma is another application of the envelope theorem, this time for constrained optimization.
- Consider the following constrained maximization problem
  \[ M(a) = \max_{x_1, x_2} g(x_1, x_2, a) \text{ such that } h(x_1, x_2, a) = 0 \]
- Setting up the lagrangian for this problem and obtaining FOC,
  \[ \frac{\partial g}{\partial x_1} - \lambda \frac{\partial h}{\partial x_1} = 0 \]
  \[ \frac{\partial g}{\partial x_2} - \lambda \frac{\partial h}{\partial x_2} = 0 \]
  \[ h(x_1, x_2, a) = 0 \]
- From these conditions, you obtain the optimal choice functions, \( x_1(a), x_2(a) \) obtaining the following identity
  \[ M(a) \equiv g(x_1(a), x_2(a)) \]
Envelope theorem for constrained optimization

- The envelope theorem says that \( \frac{dM(a)}{da} \) is equal to
  \[
  \frac{dM(a)}{da} = \left. \frac{\partial g(x_1,x_2,a)}{\partial a} \right|_{x=x(a)} - \lambda \left. \frac{\partial h(x_1,x_2,a)}{\partial a} \right|_{x=x(a)}
  \]
- In the case of cost minimization, the envelope theorem implies
  \[
  \frac{\partial c(x,w)}{\partial w_i} = \left. \frac{\partial L}{\partial w_i} \right|_{x=x(w,y)} = x_i(x(w,y)) = x_i(w, y)
  \]
  \[
  \frac{\partial c(x,y)}{\partial y} = \left. \frac{\partial L}{\partial y} \right| = \lambda
  \]
- The second implication follows also from the envelope theorem and just means that (at the optimum), the lagrange multiplier of the minimization cost is exactly the marginal cost.
Comparative statics using the cost function

Shepard’s lemma relates the cost function with the conditional factor demands. From the properties of the first, we can infer some properties for the latter.

1. From the cost function being non decreasing in factor prices follows that conditional factor demands are positive.
   \[ \frac{\partial c(w,y)}{\partial w_i} = x_i(w, y) \geq 0 \]

2. From \( c(w, y) \) being HD1, it follows that \( x_i(w, y) \) are HD0

3. From the concavity of \( c(w, y) \) it follows that its hessian is negative semi-definite. From Shepard’s lemma it follows that the substitution matrix for the conditional factor demands is equal to the Hessian of the cost function. Thus,
   1. The cross price effects are symmetric.
   \[ \frac{\partial x_i}{\partial w_j} = \frac{\partial^2 c}{\partial w_i \partial w_j} = \frac{\partial^2 c}{\partial w_j \partial w_i} = \frac{\partial x_j}{\partial w_i} \]
   2. Own price effects are non-positive.
   \[ \frac{\partial x_i}{\partial w_i} = \frac{\partial^2 c}{\partial w_i^2} \leq 0 \]
   3. The vectors of own factor demands move “opposite” to the vector of changes of factor prices.
   \[ dw dx \leq 0 \]
Comparative statics using the cost function

Shepard’s lemma relates the cost function with the conditional factor demands. From the properties of the first, we can infer some properties for the latter.

1. From the cost function being non-decreasing in factor prices follows that conditional factor demands are positive.
\[ \frac{\partial c(w,y)}{\partial w_i} = x_i(w, y) \geq 0 \]

2. From \( c(w, y) \) being HD1, it follows that \( x_i(w, y) \) are HD0

3. From the concavity of \( c(w, y) \) it follows that its hessian is negative semi-definite. From Shepard’s lemma it follows that the substitution matrix for the conditional factor demands is equal to the Hessian of the cost function. Thus,

1. The cross price effects are symmetric.
\[ \frac{\partial x_i}{\partial w_j} = \frac{\partial^2 c}{\partial w_i \partial w_j} = \frac{\partial^2 c}{\partial w_j \partial w_i} = \frac{\partial x_j}{\partial w_i} \]

2. Own price effects are non-positive.
\[ \frac{\partial x_i}{\partial w_i} = \frac{\partial^2 c}{\partial w_i^2} \leq 0 \]

3. The vectors of own factor demands moves “opposite” to the vector of changes of factor prices.
\[ dwdx \leq 0 \]
Comparative statics using the cost function

Shepard’s lemma relates the cost function with the conditional factor demands. From the properties of the first, we can infer some properties for the latter.

1. From the cost function being non decreasing in factor prices follows that conditional factor demands are positive.
   \[ \frac{\partial c(w,y)}{\partial w_i} = x_i(w, y) \geq 0 \]

2. From \( c(w, y) \) being HD1, it follows that \( x_i(w, y) \) are HD0

3. From the concavity of \( c(w, y) \) if follows that its hessian is negative semi-definite. From Shepard’s lemma it follows that the substitution matrix for the conditional factor demands is equal to the Hessian of the cost function. Thus,
   \[ \begin{align*}
   &1. \text{ The cross price effects are symmetric. } \frac{\partial x_i}{\partial w_j} = \frac{\partial^2 c}{\partial w_i \partial w_j} = \frac{\partial^2 c}{\partial w_j \partial w_i} = \frac{\partial x_j}{\partial w_i} \\
   &2. \text{ Own price effects are non-positive. } \frac{\partial x_i}{\partial w_i} = \frac{\partial^2 c}{\partial w_i^2} \leq 0 \\
   &3. \text{ The vectors of own factor demands moves “opposite” to the vector of changes of factor prices. } \text{dwdx} \leq 0
   \end{align*} \]
Comparative statics using the cost function

Shepard’s lemma relates the cost function with the conditional factor demands. From the properties of the first, we can infer some properties for the latter.

1. From the cost function being non-decreasing in factor prices follows that conditional factor demands are positive.
   \[
   \frac{\partial c(w, y)}{\partial w_i} = x_i(w, y) \geq 0
   \]

2. From \( c(w, y) \) being HD1, it follows that \( x_i(w, y) \) are HD0

3. From the concavity of \( c(w, y) \) if follows that its hessian is negative semi-definite. From Shepard’s lemma it follows that the substitution matrix for the conditional factor demands is equal to the Hessian of the cost function. Thus,
   1. The cross price effects are symmetric.
      \[
      \frac{\partial x_i}{\partial w_j} = \frac{\partial^2 c}{\partial w_i \partial w_j} = \frac{\partial^2 c}{\partial w_j \partial w_i} = \frac{\partial x_j}{\partial w_i}
      \]
   2. Own price effects are non-positive.
      \[
      \frac{\partial x_i}{\partial w_i} = \frac{\partial^2 c}{\partial w_i^2} \leq 0
      \]
   3. The vectors of own factor demands moves “opposite” to the vector of changes of factor prices. \( dwdx \leq 0 \)
Cost Minimization Second Order Conditions Conditional factor demand functions The cost function Average and Marginal Costs Geometry of Costs

Average and Marginal Costs

- **Cost Function:** \( c(w, y) \equiv wx(w, y) \)

Let’s break up the \( w \) in fixed \( w_f \) and variable inputs \( w_v \), \( w = (w_f, w_v) \). Fixed inputs enter optimization as constants.

- **Short-run Cost Function:** \( c(w, y, x_f) = w_v x_v(w, y, x_f) + w_f x_f \)

- **Short-run Average Cost:** \( SAC = \frac{c(w, y, x_f)}{y} \)

- **Short-run Average Variable Cost:** \( SAVC = \frac{w_v x_v(w, y, x_f)}{y} \)

- **Short-run Average Fixed Cost:** \( SAFC = \frac{w_f x_f}{y} \)

- **Short-run Marginal Cost:** \( \frac{\partial c(w, y, x_f)}{\partial y} \)

- **Long-run Average Cost:** \( LAC = \frac{c(w, y)}{y} \)

- **Long-run Marginal Cost:** \( LMC = \frac{\partial c(w, y)}{\partial y} \)
The total cost function is usually assumed to be monotonic, the more we produce, the higher the costs.

The average cost could be increasing/decreasing with output. We generally assume it achieves a minimum. The economic rationale is given by:

1. Average variable cost could be decreasing in some range. Eventually, it would become increasing.
2. Even if they are increasing all the way, average fixed costs are decreasing.

The minimum of the average cost function is called the **minimal efficient scale**.

Marginal cost curve cuts SAC and SAVC from the bottom at the minimum.
• Constant Returns to Scale: If the production function exhibits constant returns to scale, the cost function may be written as $c(w, y) = yc(w, 1)$

• Proof: Assume $x^*$ solves the cost minimization problem for $(1, w)$ but $yx^*$ does not the problem for $(y, w)$. Then there is $x'$ that minimizes costs for production level $y$ and $w$, and it is different than $x$. This implies that $wx' < wx$ for all $x$ in $V(y)$ and $f(x') = y$. But then, CRS implies that $f(x'/y) = 1$, so $x'/y \in V(1)$ and $wx'/y < w\bar{x}$ for all $\bar{x} \in V(1)$. We found a input vector $x'$ that is feasible at $(w,1)$ and it is cheapest way of producing it. This means that $x^*$ is not cost-minimizing. From this proof follows that $c(w, y) = wx(w, y) = wyx(w, 1) = ywx(w, 1) = yc(w, 1)$. 
For the first unit produced, marginal cost equals average cost

Long run cost is always lower than short run cost

Short run marginal cost should be equal to the long run marginal costs at the optimum (application of the envelope theorem). If $c(y) \equiv c(y, z(y))$ and let $z^*$ be an optimal choice given $y^*$ †

Differentiation with respect to $y$ yields,

$$\frac{dc(y^*, z(y^*))}{dy} = \frac{\partial c(y^*, z^*)}{\partial y} + \frac{\partial c(y^*, z^*)}{\partial z} \frac{\partial z(y^*)}{\partial y} = \frac{\partial c(y^*, z^*)}{\partial y}$$

Where the last inequality follows from FOC of the minimization problem with respect to $z$ ($\frac{\partial c(y^*, z^*)}{\partial z} = 0$)