The Use of Computational General Equilibrium Models in Tax Analysis

Nicolaus Tideman

A general equilibrium model is a description of an economy in terms of supply and demand relationships. A computable general equilibrium (CGE) model is a general equilibrium model that is solved numerically. The way that CGE models are used is to identify a predicted future path of an economy with existing taxes, change the model to reflect an alternative pattern of taxes, and see how that affects the path of the economy—wages, capital accumulation, income levels, etc.

A CGE model's level of complexity depends on the ambition of the modeler and on what data are available. The modeler must first determine the level of aggregation. The most aggregate model uses a single production function for a single commodity that can be used either for consumption or as capital. The inputs into the production function are labor, capital, and possibly land. There is a single utility function for a single type of consumer, a single place where all production and consumption occur, and the model is solved for a single moment in time.

A more elaborate model may describe a variety of economic sectors, each with its own production function and its use of products of other sectors. The inputs to the production function can be various types of labor and such capital categories as machines, structures, minerals, and infrastructure. The amount of output can depend on location as well as on the quantity of land. Consumers can vary in their age, family structure, wealth, wages, relative concern for present and future consumption, and perhaps other characteristics as well. There can be many regions, with a matrix of transport costs to determine when it is worth sending goods from one region to another.

Some inputs (land or structures, for example) may be immobile, while other inputs (people or capital, for example) are allowed to move from one region (country) to another. Production and consumption may occur in several time periods, with the amount of saving determining additions to the stock of capital.

Economic theorists tend to work with production functions and utility functions that have no specific functional form, but the creators of CGE models must specify functional forms to be able to obtain a numerical solution for the model's equilibrium. One of the simplest functional forms is the Cobb-Douglas production function,

$$Q = \phi L^{\alpha} K^{(1-\alpha)}, \tag{1}$$

where Q is the output of an economy, L is the quantity of labor used in the economy, K is the quantity of capital used in the economy, and ϕ and α are parameters that are used to calibrate the data to an economy. The parameter α specifies the share of output that goes to labor when wages equal the marginal product of labor. The parameter ϕ reflects the overall productivity of an economy. When data show that economic productivity increases persistently over time, the parameter a is often replaced by the expression $a(1+g)^t$, where g is the annual rate of autonomous productivity growth and t is the number of years from the beginning of the period being described. Then the production function is

$$Q = \phi(1+g)^t L^{\alpha} K^{(1-\alpha)}. \tag{2}$$

The exponents on labor and capital must sum to one to ensure that the production function is "homogeneous of degree one." This means that a 1% increase in all inputs leads to a 1% increase in output. When all factors are paid according to their marginal products, a production function must have the characteristic, in the vicinity of its equilibrium, that a 1% increase in all inputs leads to a 1% increase in output, to ensure that the value of the output equals the payments to the factors of production—that is, to ensure there is neither an economy-wide loss nor an economy-wide profit. Even if actual production functions are not always homogeneous of degree one, they need to have the characteristic of having a 1% increase in all scarce inputs produce a 1% increase in output, in the vicinity of the actual input combination, in order for there to be a competitive equilibrium.

There is no limit to the number of factors of production for Cobb-Douglas production function. If land (T) is included as a separate factor of production, then the function becomes

$$Q = a(1+g)^t L^{\alpha} T^{\beta} K^{(1-\alpha-\beta)}, \tag{3}$$

with the exponents again summing to 1.

Whether a two-factor model suffices or whether capital and land ought to be modeled as two different factors depends on whether the data that are used to calibrate the model contain instances of land and capital being used in significantly different proportions, with the ratio of their marginal products changing as their proportions change. If this is not the case for the available data, then the economy can be accurately described with land and capital combined into a single factor. If it does happen, the two factors must be considered separately to model the economy accurately.

The Cobb-Douglas production function has the characteristic that, no matter how much the ratio of capital to labor changes, the percentage of output that is paid to labor remains constant. Seeing this as an unfortunate limitation, economists developed the more general "constant elasticity of substitution," or CES, production function, specified by

$$Q = (1+g)^t(a_1L^{\alpha} + a_2T^{\alpha} + a_3K^{\alpha})^{1/\alpha}. \quad (4)$$

The final exponent $1/\alpha$ insures that the function is homogeneous of degree 1.

In the CES function, the share of output that is paid to labor is given by

$$\frac{a_1L^{\alpha}}{a_1L^{\alpha} + a_2T^{\alpha} + a_3K^{\alpha}}, \quad (5)$$

which varies as the factor proportions vary. The "elasticity of substitution" from which the CES function gets its name is the percentage change in the ratio of the marginal products of two factor prices that results from a 1% change in the ratio of their quantities. A Cobb-Douglas production function always has an elasticity of substitution of 1. Because a factor's share of output is constant, a 1% increase in its relative quantity is always associated with a 1% decrease in its marginal product. For the CES production function, the elasticity of substitution between any two factors is always $1/(1-\alpha)$. Thus the elasticity of substitution in the CES function can take any positive value except 1, which would require a value of α of 0 and therefore a final exponent of infinity. The CES function approaches the Cobb-Douglas function in the limit as α approaches 0. Most CGE modeling is done with CES production functions.

If the CES production function is too restrictive to describe the available data, then the step in generality that is usually next is to use the "translog" production function. The form of this production function cannot sensibly be applied to the whole space of possible combinations of factors; it is intended instead as a reasonable approximation over a range of input combinations that are to be studied. The idea behind the translog production function is that whatever transcendental form a production function may have, it is possible to approximate this form by expressing the logarithm of output as a quadratic function of the logarithms of all inputs. This functional form requires the modeler to specify how the marginal product of each factor varies with changes in every other factor.

The second key component of a CGE model is the utility function. In an aggregated model, the arguments of the utility function are the amounts of goods and leisure that a person consumes over time and within in each time period. The customary form for the utility function is "nested CES." That is, the measure of the consumption that a person has in a given year is a CES function of goods, G, and leisure, Z:

$$C_t = (b_1 G_t^{\beta} + b_2 Z_t^{\beta})^{1/\beta}. \quad (6)$$

(If there are several types of goods in the model, there are correspondingly more terms in the function for the index of consumption.) The final exponent of $1/\beta$ is not as necessary as with a production function, since utility is defined only up to a monotonic transformation, but it is somewhat useful to have the index of consumption double when all arguments double, and in the case of a negative β, the final exponent serves to ensure that utility is an increasing function of its arguments.

The next step is to specify the way in which utility depends on indexes of consumption in different time periods. Some CGE models treat the consumers in the model as families that live forever, while others treat them as persons with finite lives. For CGE models of persons with finite lives, sometimes the life is treated as having a specified length, and sometimes the life is treated as ending at a time determined by a random process. In all cases, it is customary to treat each individual as maximizing an intertemporal utility function of a CES form. For the case of infinitely lived individuals, the function is

$$U = \left[\sum_{t=0}^{\infty} \left(\frac{1}{1+\rho} \right)^{t} C_t^{\gamma} \right]^{1/\gamma} . \tag{7}$$

In this expression, ρ is known as the "pure rate of time preference," which describes what the interest rate would be if consumption were the same from one year to the next. The "intertemporal elasticity of substitution" that relates changes in the ration of consumption in different time periods to changes in the relative price of consumption in different time periods, is connected to γ in the same way that the elasticity of substitution in production is related to α. The equilibrium solution of utility function (7) shows that individuals choose that pattern of consumption at which the ratio of the price of consumption (taking account of investment opportunities) to the marginal utility of consumption is the same in all periods.

When individuals are modeled as having finite lives, the label of infinity in the utility function is replaced by the number of years between the beginning of adulthood and death, and some new individuals start their consumption each year. Such models are known as "overlapping-generations" models.

The amount of labor that the economy uses is determined by the maximization of individual utility functions. Individuals supply that amount of labor at which the marginal disutility from the loss of leisure of the last hour supplied equals the marginal utility from the goods that the individual can buy with the wages earned in the last hour worked.

While idle labor provides utility in the form of leisure, idle capital does not serve any purpose. CGE models therefore solve for the equilibrium price at which the entire capital is used productively. The capital stock in any year is the previous year's stock, minus last year's depreciation, plus the capital purchased with the individual income that did not go to taxes or consumption.

The descriptions of labor and capital use in equilibrium presume that the economy is "closed," so that there is no opportunity for labor or capital to move in or out of the economy. In an open economy, the wage and the interest rate are determined as the global equilibrium, so any local economy employs labor up to the point at which the value of the labor's marginal product falls below the global wage, and it employs the amount of capital that is able to yield the rate of return determined by the more global economy.

If land is specified as a factor of production in the model, then the usual assumption is that the quantity of land use in the economy is fixed. Because land cannot migrate across regions, each region can be expected to have a different equilibrium rent of land .

To analyze the effects of taxation, it is necessary to adjust the equations that describe the decisions of firms and consumers to incorporate the distortive effects of different taxes. For example, income taxes reduce the return to work and to saving, sales taxes increase the cost of consumption, and property taxes and corporation income taxes reduce the return to investment. The model is used to predict how changes in taxes would affect the future path of the economy.

The strategies for solving a CGE model are based on the idea that, if the economic structure of an economy is adequately specified, then there will be as many equations as there are variables to be solved for. When that equality prevails, there is generally a unique solution. Multiple equilibria are possible in principle, but unlikely. CGE models are often complex enough to require substantial computer resources and substantial time to solve. Therefore the modeler must be economical in what is asked of a model.

There are two principal methods by which CGE models are solved, the Newton-Raphson method and the simplex method. The Newton-Raphson method requires the modeler to start with guesses for all of the variables whose equilibrium values must be determined to solve the model. The solution algorithm then makes calculations that are intended to determine the magnitudes of all disequilibria in the economy with the specified values. It makes slight changes in the variables, one by one, to determine all of the first and second derivatives of the disequilibria with respect to the variables. If the second derivatives were all constant, then it would be a simple matter to compute the equilibrium vector at which all disequilibria vanish. Since the second derivatives are not constant, the modeler settles for moving some fraction of the computed distance, in the direction of the computed equilibrium, and then recomputes the derivatives to make a new estimate of where to go. There is no guarantee that this process will converge to an equilibrium, but if it does converge, then the process will usually be very fast.

The alternative simplex method does not require the modeler to guess good starting values and guarantees convergence to the equilibrium. The "simplex" refers to the *n*-dimensional generalization of a triangle. It is based on the idea that a set of relative prices can be considered a point in a simplex in which a lattice is formed, with each point in the lattice representing a different pattern of relative prices. A set of rules is used to specify how one decides which lattice to examine next, on the basis of the disequilibria that result from the current lattice point. Eventually, these rules lead to a path that repeats itself. When this happens, the lattice is replaced by a finer one, and the search continues.

The lattice nature of the simplex method provides a guarantee that the search will get within any specifiable distance of the equilibrium within a specifiable, finite number of steps. The disadvantage of the simplex method is that it is very slow in solving large models with large numbers of variables whose equilibrium values need to be determined.

Since the simplex method has its greatest advantage when the modeler has little idea where the equilibrium is while the Newton-Raphson method is most advantageous when the equilibrium is close by, the modeler may wish to start with the simplex method and shift to Newton-Raphson when the lattice has become rather fine.

One of the difficulties of using CGE models is that it is reasonable to expect people to make their plans over spans of time of indefinite duration, but a model cannot trace the path of an economy to the end of time. Some modelers confine their attention to comparing the equilibria of economies, but this is problematic because it may take generations for an economy to reach an equilibrium. What happens on the way to equilibrium is very important. Other modelers trace the paths of an economy for hundreds of years, figuring that when so much time has passed, any difficulties with respect to the way the model ends will have no effect on the initial years, which is the real concern of the model.

I discovered a different way of dealing with the indefinite planning horizon. I have found that when a model for a limited number of years is a good description of an economy, the path of the rate of saving over time will be closely approximated by one of the simple functional forms that describe a second-order differential equation with a horizontal asymptote. That is, the saving rate can be approximated by either the difference of two exponential decays or the product of an exponential decay and a sine wave. When the path of saving does not have this shape, the model is a poor description of an economy. So I look for a description of an economy that gives the right shape to the beginning of the path of the saving rate over time, and I have a reasonable approximation to the path that the beginning of the saving rate would have if I had modeled a century or more.

Summary

If you want to estimate the impact of a tax change on an economy, begin by deciding how complex a model you want, taking account of the availability of data and the programming and computing challenges of more complex models, along with the potential for better insights from more complex models. Decide what programming environment and what solution technique you want to use. Develop the model and gather the data to calibrate the model to the economy you want to approximate. Run the model with existing taxes and identify the paths of such variables of interest as wages, incomes and the accumulation of capital. Substitute alternative taxes that you believe would be better, re-run the model, and measure the changes in the variables of interest.