Economic Review

FEDERAL RESERVE BANK OF DALLAS
FIRST QUARTER 1998

The *Economic Review* is published by the Federal Reserve Bank of Dallas. The views expressed are those of the authors and do not necessarily reflect the positions of the Federal Reserve Bank of Dallas or the Federal Reserve System.

Subscriptions are available free of charge. Please send requests for single-copy and multiple-copy subscriptions, back issues, and address changes to the Public Affairs Department, Federal Reserve Bank of Dallas, P.O. Box 655906, Dallas, TX 75265-5906, (214) 922-5257.

Contents

Gasoline is the petroleum product whose price is most visible and, therefore, always under public scrutiny. Many claim there is an asymmetric relationship between gasoline and oil prices—specifically, gasoline price changes follow oil price changes more quickly when oil prices are rising than when they are falling. To explore this issue, Nathan Balke, Stephen Brown, and Mine Yücel use several different model specifications to analyze the relationship between oil prices and the spot, wholesale, and retail prices of gasoline. They find asymmetry is sensitive to model specification but is pervasive with the most general model.

The controversy over the success or failure of NAFTA is now bleeding over into discussions about the benefits of extending the trade accord to other countries in the Western Hemisphere. The NAFTA debate has typically focused on its impact on employment. But to understand the overall economic effects of NAFTA, it is important to first determine its impact on trade. In this article, David Gould explores NAFTA's effects on North America's trading patterns since its implementation in 1994. He finds that although it is difficult to distinguish any effect of NAFTA on trade between Canada and Mexico or Canada and the United States, trade between the United States and Mexico has significantly increased since 1994.

The Internal Revenue Service remains unpopular, the U.S. savings rate remains low, and pressure to efficiently raise significant new tax revenues seems certain to grow once the baby boom generation reaches retirement age. Consequently, it is likely that alternatives to the current income tax system will receive substantial political and media attention in coming years.

In this first of two articles on the economic impact of fundamental tax reform, Evan Koenig and Gregory Huffman describe a framework for analyzing how the adoption of a flat-rate consumption tax would affect the economy over time. Their analysis indicates that replacing the income tax with a consumption tax would have an immediate positive impact on saving and lead, in the long run, to higher levels of consumption, wages, and stock prices and to lower interest rates. In the short run, however, interest rates would probably rise, and consumption and stock prices would probably decline.

Crude Oil and Gasoline Prices: An Asymmetric Relationship?

Nathan S. Balke
Research Associate
Federal Reserve Bank of Dallas
and
Associate Professor of Economics
Southern Methodist University

Stephen P. A. Brown
Senior Economist and Assistant Vice President
Federal Reserve Bank of Dallas

Mine K. Yücel
Senior Economist and Research Officer
Federal Reserve Bank of Dallas

Many claim to observe an asymmetric relationship between gasoline and oil prices…that gasoline prices respond more quickly when oil prices are rising than when oil prices are falling.

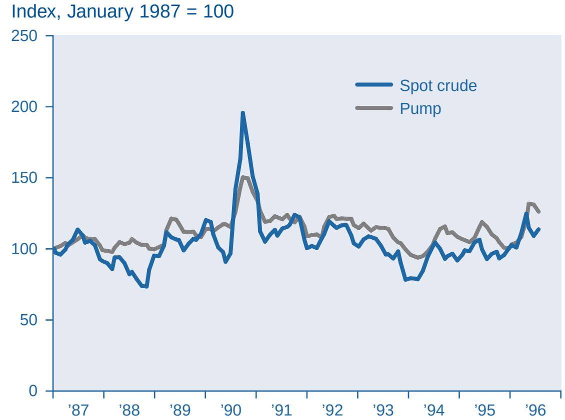
Gasoline accounts for about half the U.S. consumption of petroleum products, and its price is the most visible among these products. As such, changes in gasoline prices are always under public scrutiny. Many claim to observe an asymmetric relationship between gasoline and oil prices—specifically that gasoline prices respond more quickly when oil prices are rising than when oil prices are falling (*Figure 1*). President George Bush gave these concerns official weight during the Gulf War when he asked the oil companies to show restraint in raising prices for their products.

Much of the previous research provides econometric support for public claims of asymmetry in the movements of gasoline and crude oil prices. Bacon (1991) found asymmetry in the U.K. gasoline market, and Karrenbrock (1991), French (1991), Borenstein, Cameron, and Gilbert (1997), and a GAO report (1993) all found some evidence of an asymmetric response in U.S. gasoline markets. Norman and Shin (1991) found a symmetric response in U.S. gasoline markets.

Of these studies, one of the most comprehensive and compelling is that of Borenstein, Cameron, and Gilbert (1997), hereafter identified as BCG. They use a series of bivariate error-correction models to test for asymmetry in price movements in each of the various stages in the production and distribution of gasoline from the crude oil price through the refinery to the retail pump, using weekly and biweekly data from 1986 to 1992. They find strong and pervasive evidence of asymmetry.

As Shin (1992) has argued, however, the periodicity of the data, the sample period of estimation, and the model specification may have affected the results obtained in previous studies. To explore these issues, we extend the work of BCG by using several different model

Figure 1
Spot Crude Oil Prices and Retail Gasoline Prices
Index, January 1987 = 100



specifications with weekly data from 1987 through early 1996. We find that most of the price volatility originates upstream. We also find econometric evidence of asymmetry in the extended sample. The findings are sensitive to model specification but not to sample period.

Although popular opinion attributes asymmetry to market power, a number of competing explanations have been offered. This article presents an econometric exercise that attempts to identify whether asymmetry occurs; it does not address how asymmetry might arise. For an overview of the possible explanations, see the box entitled, "Why Does Asymmetry Arise?"

THE ORIGINATION AND TRANSMISSION OF PRICE SHOCKS

In theory, price shocks can originate at any point from crude oil prices to the final price at the gasoline pump. Shocks originating at an intermediate step, such as the wholesale price of gasoline, may reflect a bottleneck in distribution, while price shocks originating farther upstream are more likely to represent the effects of variation in crude oil supply. Price shocks originating at the retail level are more likely to represent variation in U.S. demand for gasoline. Given the history of oil-supply shocks and indications that demand for gasoline is relatively stable, intuition suggests that price shocks are more likely to originate upstream and be transmitted downstream.

To examine where price shocks originate and how they are transmitted across the U.S. market for gasoline, we use time-series methods. Specifically, we test for Granger causality and compute variance decompositions for each pair of upstream and downstream prices, including spot crude oil prices and spot, wholesale, and retail gasoline prices. As reported below, both the causality tests and variance decompositions generally confirm the intuition that price shocks more frequently begin upstream and are then transmitted downstream.

To motivate the relationship between a pair of upstream and downstream prices, consider a simple markup model,

$$(1) \qquad PD_t = a + bPU_t ,$$

where PD_t is a downstream price, PU_t is an upstream price, and a and b are parameters indicating the relationship between the upstream and downstream prices.[1] The markup, a, represents the cost of refining, marketing, transportation, and/or distribution. The scalar, b, allows for differences in units and heat content but may also reflect other market phenomena.

The time-series analysis involves several steps. We check whether the prices are stationary. We then test for Granger causality, which allows an assessment of the lead–lag relationship between each pair of prices. Finally, we calculate the variance decompositions to assess the sources of shocks to the variables.

Data

To analyze the relationships between crude oil and gasoline prices, we use weekly data from January 1987 through August 1996. The oil price is the spot price for West Texas Intermediate crude, the spot price for gasoline is the New York Harbor Spot Price for unleaded regular, and the retail price is the self-service pump price for regular unleaded motor gasoline, with and without taxes. These series are obtained from the Weekly Energy Statistics of Haver Analytics. The wholesale price is from the Oil Price Information Service and represents an average wholesale price for unleaded gasoline across all U.S. wholesale distributors reporting data continuously from 1986 through August 1996.

Stationarity

As an initial step in our econometric work, we perform several diagnostic checks to assess the correct specification for the various series. We test for nonstationarity using augmented Dickey–Fuller and Phillips–Perron tests and conclude that we can reject the hypothesis that the series have a unit root. Because all our price series appear to be stationary, we represent the relationship between any pair of prices in log levels.

Causality

A causal relationship between two variables implies that changes in one variable lead changes in the other. To assess the lead–lag relationships for each pair of variables, we perform bidirectional Granger causality tests on each of the ten pairs of upstream and downstream prices as follows:

$$(2)\ PD_t = \alpha_1 + \sum_{i=1}^{m} \beta_{1,i} PU_{t-i} + \sum_{i=1}^{m} \delta_{1,i} PD_{t-i} + \mu_{1,t} \text{ and}$$

$$(3)\quad PU_t = \alpha_2 + \sum_{i=1}^{n} \beta_{2,i} PD_{t-i} + \sum_{i=1}^{n} \delta_{2,i} PU_{t-i} + \mu_{2,t},$$

where PD_t is the downstream price; PU_t is the upstream price; α_1, $\beta_{1,i}$, $\delta_{1,i}$, α_2, $\beta_{2,i}$, and $\delta_{2,i}$ are parameters to be estimated; and $\mu_{1,t}$ and $\mu_{2,t}$ are white-noise residuals. The lag length used for estimation of each equation is the shortest lag length that yields white-noise residuals (as indicated by the Ljung–Box Q statistic with a probability of 10 percent).

Why Does Asymmetry Arise?

With a number of studies showing that gasoline prices respond more quickly when crude oil prices rise than when they fall, analysts have offered a number of explanations for the phenomenon.[1] Explanations include market power, search costs, consumer response to changing prices, inventory management, accounting practices, and refinery adjustment costs. For the banking industry, Neumark and Sharpe (1992) show that market concentration is an explanatory variable for the asymmetry found in interest rate movements. For the gasoline markets, however, no one has posited econometric tests that would allow the testing of the various explanations (including market power) for price asymmetry against the available data. Without such tests, it remains a matter of speculation whether the asymmetric response of gasoline prices to movements in crude oil prices is the result of market power or more benign forces.

Market Power and Search Costs

Market power is probably the greatest concern to those who observe that gasoline prices respond more quickly when crude oil prices rise than when they fall. Yet no formal model shows a relationship between market structure and asymmetric response of downstream prices to changes in upstream prices.[2] Were such a model to exist, it might involve firms that are concerned with maintaining a tacit collusion and/or consumer search costs.

Consider an industry with a few dominant firms that are engaged in an unspoken collusion to maintain higher profit margins. Reputation can be important to maintaining such a tacit agreement (Tirole 1990). If the firms value the tacit agreement and have imperfect knowledge of the upstream prices its competitors are paying, then each firm would face an asymmetric loss function in which it would be more reluctant to lower its selling price than to raise it. When upstream prices rise, each firm is quick to raise its selling prices because it wants to signal its competitors that it is adhering to the tacit agreement by not cutting its margin. When the upstream price falls, each firm is slow to lower its selling price because, in doing so, it runs the risk of sending a signal to its competitors that it is cutting its margin and no longer adhering to the tacit agreement. In the gasoline markets, such an explanation could be applied to each upstream price and its adjacent downstream price.

In the retail gasoline market, consumer search costs could lead to temporary market power for gas stations and an asymmetric response to changes in the wholesale price of gasoline. (See BCG, Norman and Shin 1991, Borenstein 1991, and Deltas 1997.) Each gas station has a locational monopoly that is limited only by consumer search. After consumers have searched, the profit margins at each gas station are pushed down to a roughly competitive level. When wholesale prices rise, each station acts to maintain its profit margins and quickly passes the increase on to customers. When wholesale prices fall, however, each station temporarily boosts its profit margins by slowly passing the decrease on to customers. Only after the customers engage in a costly and time-consuming search to find the lowest prices are the stations forced to lower prices to a competitive level.

More Benign Explanations

Although the existence of asymmetry could be consistent with market power, it is not the only explanation that economists have offered for the asymmetric response of gasoline prices to movements in crude oil prices. Alternative explanations include consumer response to changing prices, inventory management, accounting practices, and refinery adjustment costs.

An asymmetric consumer response to changing gasoline prices may contribute to the asymmetry between movements in crude oil and gasoline prices. If consumers accelerate their gasoline purchases to beat further increases when its price is rising, they will increase inventories held in automobiles and quicken the pace at which the price rises. If consumers fear running out of gasoline and do not slow their purchases of it when its price is falling by as much as they accelerated their purchases when prices rose, then the price of gasoline will fall more slowly than it rose.

Similarly, firms in the oil industry may view the short-run costs of unexpected changes in their inventories as asymmetric. (See BCG.) If the costs of operations rise sharply when inventories are reduced below normal operating levels, a reduction of upstream supply could lead a firm to raise its output prices aggressively to prevent a loss of inventories. If an increase in inventories above normal operating levels has a relatively small effect on costs, the firm could be less aggressive in reducing its selling prices when it experiences an increase in upstream supply. Hence, inventories would buffer downstream price movements less when prices are rising than when they are falling.

The asymmetry arising from changes in inventories could be enhanced by FIFO (first in, first out) accounting. If inventories fall when upstream supply is reduced, the firm will sell the products incorporating the higher upstream price sooner. If inventories rise when upstream supply is increased, the firm will sell the products incorporating the lower upstream price later.

Refiners also face high adjustment costs to changing their output, and, when possible, they slowly adjust output. When crude oil supplies are reduced and inventory reductions are costly, however, refiners as a group have little choice but to reduce output quickly, which would lead to fairly quick increases in gasoline prices. When crude oil supplies are increased, however, refiners slowly increase output, delaying the decreases in gasoline prices.

Notes

[1] Pricing asymmetries have been observed in a number of industries, including banking (Neumark and Sharpe 1992) and agriculture (Mohanty, Peterson, Wesley, and Kruse 1995).

[2] Variations of the kinked-demand model of oligopoly do not suggest an asymmetrical movement in the output price of an industry in response to common shocks to the input prices of the firms in that industry. See Scherer (1980) and Neumark and Sharpe (1992).

Causality runs from the upstream price to the downstream price if the coefficients $\beta_{1,i}$ are jointly significantly different from zero. Similarly, causality runs from the downstream price to the upstream price if the coefficients $\beta_{2,i}$ are jointly significantly different from zero.

In most cases, upstream prices seem to contain market information that is later incorporated in the downstream prices. As Table 1 shows, we find that causality runs from the upstream price to the downstream price for each pair, with two exceptions. The spot price for crude oil does not appear to lead the spot price for gasoline, nor does the retail price of gasoline without taxes seem to lead the retail price of gasoline including taxes. We do find, however, that the spot price for gasoline leads the spot price for crude oil, and the retail price including taxes leads the retail price excluding taxes. These findings suggest the possibility that for these two pairs of prices, information is incorporated in the downstream price a bit more quickly than in the upstream price. We also find that each of the gasoline prices Granger-cause the spot price for crude oil, which suggests that each of these prices contains market information that is later incorporated into the spot price for crude oil.[2]

Long-Run Sources of Variance

To find out which price shocks have been sources of volatility during the sample period, we construct a bivariate vector autoregressive (VAR) model to represent each relationship and calculate the variance decomposition for each pair of prices.[3] For given time horizons, the variance decomposition apportions the stochastic variability in a given price to shocks originating in itself and to shocks originating in the price with which it is paired. We consider a 30-week time horizon, which should represent the long run because the variance decomposition shows a minimal change after 30 weeks.

As Table 2 shows, the variance decompositions generally suggest that over the long run, price shocks originate in upstream prices and are transmitted downstream. In addition, the variance decompositions suggest that proximity enhances the importance of the upstream price as a source of variation in a downstream price. The one exception is in the relationship between the spot price for crude oil and the spot price for gasoline. In its pairing with the spot price of gasoline, the spot price for crude oil accounts for about one-half of the variance in the gasoline price over the long run. In its more distant pairing with the wholesale price of gasoline, the spot price of crude oil accounts for about two-thirds of the variance in the gasoline price over the long run.

Table 1

Significance of Granger Causality Tests

(Variable at left is the dependent variable)

	Oil	Spot gasoline	Wholesale	Retail without tax	Retail with tax
Oil	—	.014	.015	.024	.049
Spot gasoline	.53	—	.176	.097	.229
Wholesale	.002	.0	—	.455	.088
Retail without tax	.0	.0	.0	—	.0
Retail with tax	.0	.0	.0	.0	—

Table 2

Decomposition of Variance

(Percentage of forecast error variance of dependent variable explained by shocks to independent variable)

	Oil	Spot gasoline	Wholesale	Retail without tax	Retail with tax
Oil	—	18.1	2.7	4.4	.21
Spot gasoline	48.9	—	1.8	.47	.65
Wholesale	63.5	85.8	—	.15	.53
Retail without tax	63.3	84.7	87.2	—	.47
Retail with tax	45.8	42.8	83.7	91.5	—

NOTE: The variance decompositions are from bivariate VARs. The variable at left is the dependent variable. The pair orderings are from upstream to downstream.

A BASIC MODEL OF ASYMMETRY

Given the findings that price volatility most often originates in the upstream price of any price pair and that causality is stronger going from upstream prices to downstream prices, we restrict our inquiry to those cases in which the downstream price is the dependent variable and the upstream price is the independent variable, as is suggested by Equation 1. The relationships between upstream and downstream prices, coupled with the finding that each of the variables is stationary, suggest modeling asymmetry in levels as follows:

$$(4)\quad PD_t = \alpha + \sum_{i=0}^{n} \beta_i PU_{t-i} + \sum_{i=1}^{n} \gamma_i PD_{t-i} + \sum_{i=0}^{n} \delta_i U_{t-i} PU_{t-i} + \sum_{i=1}^{n} \lambda_i D_{t-i} PD_{t-i} + \mu_t,$$

where U_{t-i} is a variable that takes a value of one when PU_{t-i} is greater than PU_{t-i-1} and is zero otherwise; D_{t-i} is a variable that takes a value of one when PD_{t-i} is greater than PD_{t-i-1} and is zero otherwise; α, β_i, γ_i, δ_i, and λ_i, are parameters to be estimated; and μ_t is a white-noise

Table 3
Significance of Asymmetry Tests for Levels Model
(Variable at left is the dependent variable)

	Asymmetry type	Oil	Spot gasoline	Wholesale	Retail without tax
Spot gasoline	Indep. Var.	.075	—	—	—
	Total	.109	—	—	—
Wholesale	Indep. Var.	.239	.011	—	—
	Total	.097	.0	—	—
Retail without tax	Indep. Var.	.0	.42	.08	—
	Total	.0	.24	.013	—
Retail with tax	Indep. Var.	.0	.89	.35	.15
	Total	.0	.048	.067	.28

residual.[4] To facilitate comparison with BCG and to control for seasonal and time-varying pricing patterns, we include 51 weekly dummies and a time-trend variable in each regression.[5] The lag length used for estimation is the shortest lag length that yields white-noise residuals.

The regression's specification allows for asymmetry in the response of the downstream price to arise either from its own history or from the upstream price. Asymmetry is indicated if the coefficients δ_i and λ_i are jointly significantly different from zero.

Table 3 indicates that symmetry is rejected in half the price pairs at the 5 percent significance level, but the results are not very systematic. For instance, the tests indicate that retail prices for gasoline, both with and without taxes, respond asymmetrically to crude oil prices, while the retail price with taxes responds asymmetrically to the spot price for gasoline, but the retail price without taxes does not. In contrast, the retail price of gasoline without taxes responds asymmetrically to the wholesale price, but the retail price with taxes does not. In two of the pairings—spot gasoline with retail including taxes and wholesale with retail without taxes—the asymmetry seems to arise from the dependent variable's own dynamics. The lack of consistent results makes it difficult to determine in which stages of the market asymmetry arises. In addition, dynamic simulations indicate that for those cases in which asymmetry is statistically significant, it is relatively small.

These findings contrast with those of BCG, who find pervasive evidence of asymmetry that is large in magnitude, using data from 1986 to 1992. We use a shorter sample, 1987–92, and find it has no effect on the results.

AN ALTERNATE SPECIFICATION

Because the sample period used for estimation does not seem to explain the difference between the results above and those of BCG, we consider the differences between the specification of Equation 4 and that used by BCG. A model similar in specification to that used by BCG yields substantially different results from the levels model.

Having found that the shorter data series they utilized are difference stationary, BCG uses an error-correction model similar to Equation 4 for estimation. Allowing for asymmetry, including in the error-correction process, one representation of the error-correction model is

$$(5)\quad \Delta PD_t = a + \sum_{i=0}^{n} b_i \Delta PU_{t-i} + \sum_{i=1}^{n} c_i \Delta PD_{t-i} + \sum_{i=0}^{n-1} d_i U_{t-i} \Delta PU_{t-i} + \sum_{i=1}^{n-1} f_i D_{t-i} \Delta PD_{t-i} + y(PU_{t-1} - zPD_{t-1}) + \mu_t,$$

where ΔPD_t is the first difference of PD_t, the downstream price; $\Delta \mathrm{PU}_t$ is the first difference of PU_t, the upstream price; a, b_i, c_i, d_i, f_i, and y are parameters to be estimated; z is the estimated parameter from the long-run relationship between PD_t and PU_t; and μ_t is a white-noise residual.

In estimation, however, BCG do not make use of the long-run restriction implied by the error-correction process, as the coefficients on the levels variables in their specification are left unrestricted, despite finding that their data series are difference stationary. Therefore, in the absence of asymmetry, their model would be equivalent to the levels model shown in Equation 4. Like BCG, we do not impose a long-run restriction in the estimation (which would not be supported by stationary data), but unlike BCG, we allow for asymmetry in the levels variables of the error-correction process, which allows us to rewrite Equation 5 as

$$(6)\quad PD_t = \alpha + \sum_{i=0}^{n} \beta_i PU_{t-i} + \sum_{i=1}^{n} \gamma_i PD_{t-i} + \sum_{i=0}^{n-1} \zeta_i U_{t-i} \Delta PU_{t-i} + \sum_{i=1}^{n-1} \eta_i D_{t-i} \Delta PD_{t-i} + \delta U_{t-1} PU_{t-1} + \lambda D_{t-1} PD_{t-1} + \mu_t,$$

where α, β_i, γ_i, ζ_i, η_i, δ, and λ are parameters to be estimated, and μ_t is a white-noise residual.[6] As with Equation 4, we include 51 weekly dummies and a time-trend variable in the regression. The lag length used for estimation is the shortest lag length that yields white-noise residuals. As is the case for Equation 4, the specification of Equation 6 allows for asymmetry in the response of the downstream price to arise either from its own history or from the upstream price. Asymmetry is indicated if the coefficients δ_i, λ_i, ζ, and η are jointly significantly different from zero.

Although Equation 6 differs from Equation 4 only in its specification of asymmetry, estimation with Equation 6 indicates more pervasive asymmetry.[7] As Table 4 shows, symmetry is rejected in nine of the ten price pairs. The error-correction specification barely rejects the hypothesis that the retail price of gasoline without taxes responds asymmetrically to the spot price of gasoline, but this is the only pairing in which asymmetry is not indicated. In two pairings—spot gasoline with wholesale and wholesale with retail without taxes—asymmetry seems to arise from the dependent variable's own dynamics. The pervasive asymmetry indicated by the error-correction model is consistent with the findings of BCG. Use of a shorter sample period, 1987–92, does not significantly affect the results.

The Magnitude of Asymmetry

To assess the extent of the asymmetry implied by the two models, we examine the response of the downstream price to both a permanent one-time increase in the upstream price and to a permanent one-time decrease in the upstream price. Figures 2 through 6 plot the differences between the downstream price's response to an increase and to a decrease in the upstream price.[8] The solid line in each figure represents the point estimate of the response, and the dashed lines represent a confidence band of two standard deviations.[9]

Figures 2 through 6 show that the asymmetry implied by the error-correction model is substantially different from that implied by the

Table 4
Significance of Asymmetry Tests for Error Correction Model
(Variable at left is the dependent variable)

	Asymmetry type	Oil	Spot gasoline	Wholesale	Retail without tax
Spot gasoline	Indep. Var.	.004	—	—	—
	Total	.004	—	—	—
Wholesale	Indep. Var.	.004	.08	—	—
	Total	.001	.0	—	—
Retail without tax	Indep. Var.	.0	.21	.06	—
	Total	.0	.08	.007	—
Retail with tax	Indep. Var.	.001	.035	.0	.0
	Total	.0	.0	.001	.0

Figure 2
Difference in Response of Spot Gasoline Price to Rising and Falling Price of Crude Oil

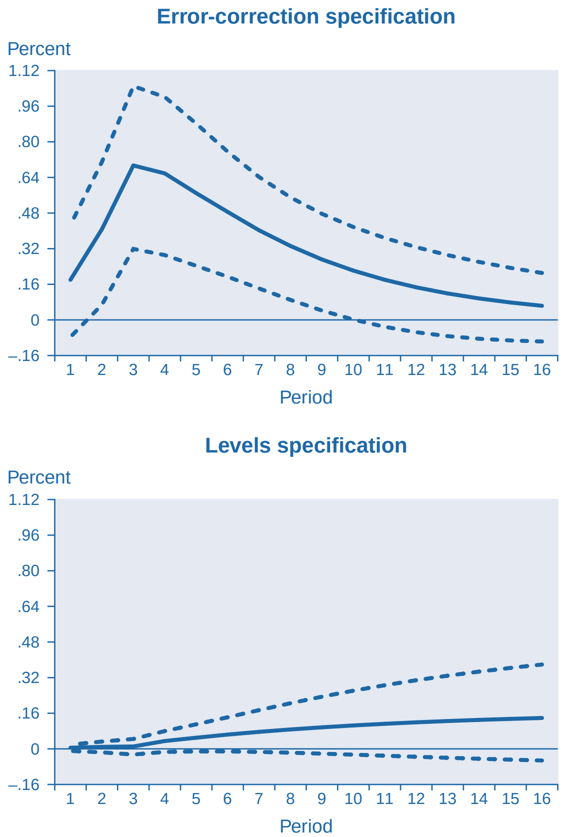


Figure 3
Difference in Response of Wholesale Gasoline Price to Rising and Falling Price of Crude Oil

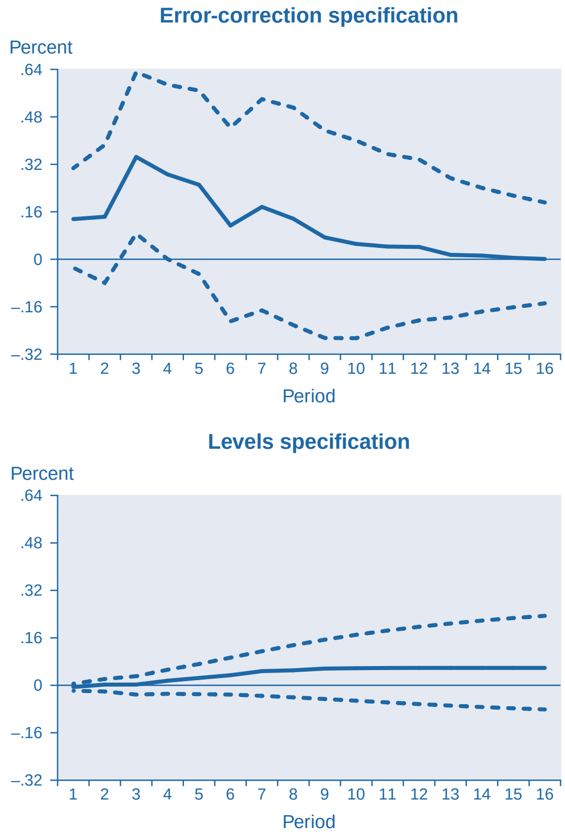


Figure 4
Difference in Response of Retail Price (With Tax) to Rising and Falling Price of Crude Oil

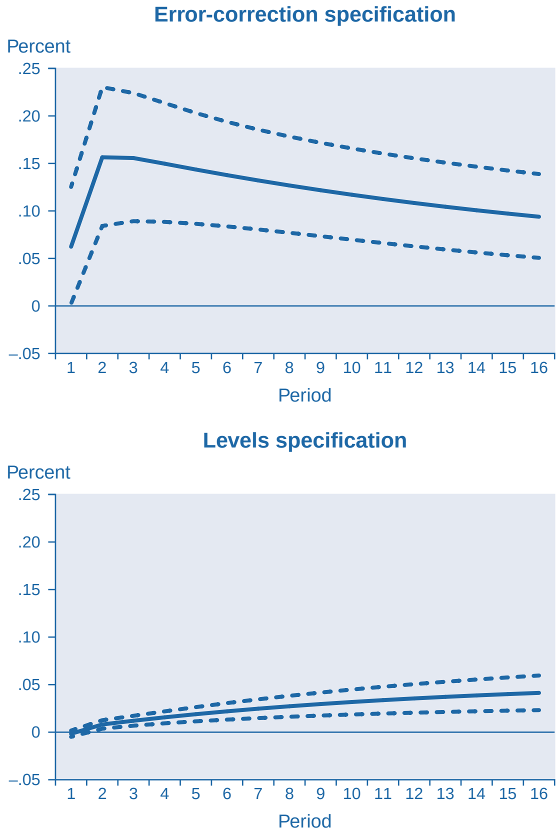
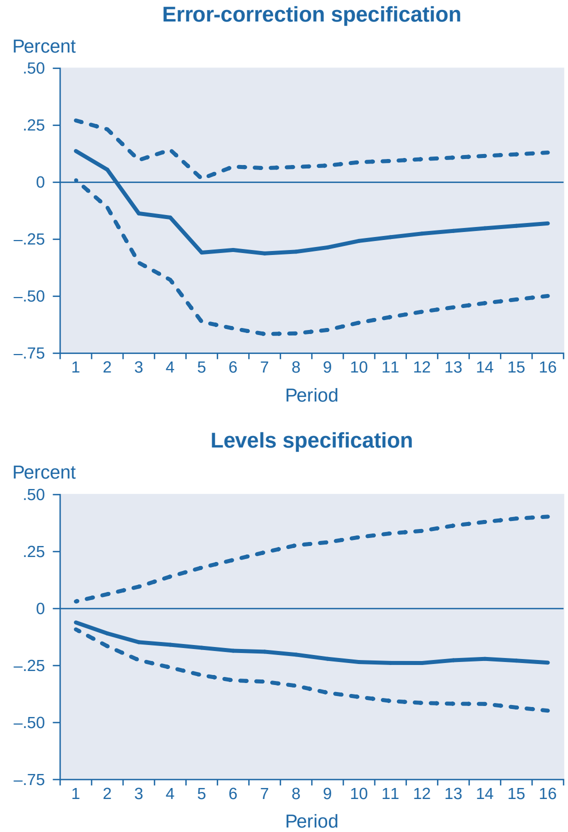


Figure 5
Difference in Response of Wholesale Gasoline Price to Rising and Falling Spot Price of Gasoline



levels model. For the error-correction model, the difference in the response of the downstream price to an increase versus a decrease in the upstream price is generally statistically significant. For the levels model, the difference is statistically significant only in a few cases. Furthermore, the magnitude of the asymmetry implied by the error-correction model is several times larger than that implied by the levels model, particularly during the first eight weeks following a change in the upstream price. Even for the price pairs in which the levels model does indicate significant asymmetry, the magnitude of the asymmetry is substantially smaller than that implied by the error-correction model (for example, see Figure 4).

For most of the error-correction models, the asymmetry peaks one or two weeks after the initial change in the upstream price and then slowly dies out. When retail (both with and without taxes) is the downstream price, the asymmetry can be fairly long-lived—longer than four months. In the few cases in which the levels model shows asymmetry, the asymmetry is quite persistent.

The one anomalous response is that of wholesale prices to changes in the spot gasoline price (*Figure 5*). In this case, both the error-correction model and the levels model imply that wholesale prices respond more to a decrease in the spot price than to an increase. The difference is not statistically significant in either model, however, in contrast to the F-tests reported in Tables 3 and 4.

DIFFERENCES IN SPECIFICATION

The fact that the two models yield such different results is puzzling. Under the null hypothesis of no asymmetry, the two models are identical (as the long-run restriction is not placed on the error-correction model). The differences arise solely in the specification of asymmetry.

To highlight the similarities and differences of the specifications represented by Equations 4 and 6, we create a generalized model in which the two specifications are nested. With some algebraic manipulation, the generalized model can be written as

$$(7)\ PD_t = a + \sum_{i=0}^{n} b_i PU_{t-i} + \sum_{i=1}^{n} c_i PD_{t-i} + \sum_{i=0}^{n} d_i U_{t-i} PU_{t-i} + \sum_{i=1}^{n} f_i D_{t-i} PD_{t-i} + \sum_{i=0}^{n-1} g_i U_{t-i} PU_{t-i-1} + \sum_{i=1}^{n-1} h_i D_{t-i} PD_{t-i-1} + \mu_t,$$

where a, b_i, c_i, d_i, f_i, g_i, and h_i are parameters to be estimated, and μ_t is a white-noise residual. As with Equations 4 and 6, we also include 51 weekly dummies and a time-trend variable. The levels specification (*Equation 4*) is obtained if the coefficients g_i and h_i are zero. The error-correction specification is obtained if $g_i = -d_i$ and $h_i = -f_i$ for all i except $i = 1$, and $d_n = 0$ and $f_n = 0$.

Unfortunately, the differences between the models do not seem to lend themselves to sharply diverging economic interpretations. Consequently, we use Equation 7 to test for asymmetry and the restrictions imposed by the two models. Asymmetry is indicated if the coefficients d_i, f_i, g_i, and h_i are jointly significantly different from zero. The restriction representing the levels specification is rejected if the coefficients g_i and h_i are jointly significantly different from zero. The restriction representing the error-correction specification is rejected if $g_i = -d_i$ and $h_i = -f_i$ for all i except $i = 1$, and $d_n = 0$ and $f_n = 0$ are jointly significantly rejected.

Table 5 shows that for eight of the ten price pairs, the restrictions implied by the levels model are rejected, but the restrictions implied by the error-correction model cannot be rejected. For one pairing—spot gasoline with retail sans tax—the restriction implied by either model cannot be rejected. In the pairing of wholesale with retail sans tax, the restriction implied by the levels model cannot be rejected, but the restriction implied by the error-correction model is rejected.

For each of the nine pairings in which one specification seems to be preferred over the other, the preferred model indicates asymmetry. For the one pairing in which neither set of restrictions can be rejected, neither model indicates asymmetry. As shown by Table 6, asymmetry tests conducted with the general model are substantially consistent with the results from the preferred model for each pairing.

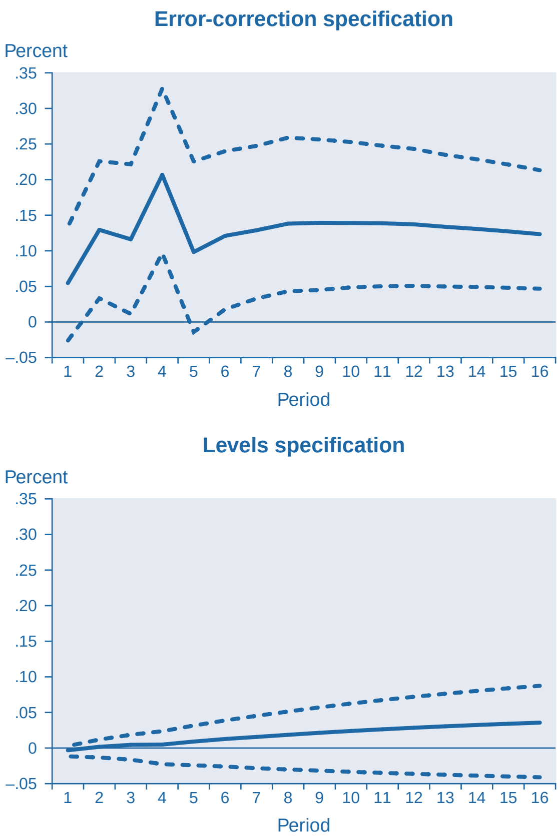
Table 5
Significance of Nesting Tests for Levels and Error-Correction Models
(Variable at left is the dependent variable)

	Asymmetry type	Oil	Spot gasoline	Wholesale	Retail without tax
Spot gasoline	LM	.01	—	—	—
	ECM	.06	—	—	—
Wholesale	LM	.006	.002	—	—
	ECM	.277	.132	—	—
Retail without tax	LM	.002	.242	.052	—
	ECM	.28	.592	.011	—
Retail with tax	LM	.0	.020	.014	.0
	ECM	.392	.301	.457	.143

Table 6
Significance of Asymmetry Tests for General Model
(Variable at left is the dependent variable)

	Asymmetry type	Oil	Spot gasoline	Wholesale	Retail without tax
Spot gasoline	Indep. Var.	.077	—	—	—
	Total	.009	—	—	—
Wholesale	Indep. Var.	.014	.013	—	—
	Total	.004	.0	—	—
Retail without tax	Indep. Var.	.001	.193	.034	—
	Total	.0	.18	.002	—
Retail with tax	Indep. Var.	.002	.061	.022	.0
	Total	.0	.002	.006	.0

Figure 6
Difference in Response of Retail Price (with Tax) to Rising and Falling Wholesale Price



Asymmetry Reconsidered

We have considered two model specifications to test for asymmetry in the response of gasoline prices to crude oil prices. Even though the two models differ only in their specification of asymmetry and are otherwise identical, they yield dramatically different results. A levels specification indicates that asymmetry is only found in a few cases and is small. An error-correction specification (without a long-run restriction) indicates that asymmetry is pervasive and large.

Unfortunately, the differences in specification do not seem to lend themselves to economic interpretation, which leaves us with a statistical criterion with which to evaluate the divergent findings. In most cases, tests with a more general model indicate that the error-correction model seems to fit the data better than the levels model, which suggests that the apparent asymmetry is one that operates on the rate of change in prices. If we accept the error-correction specification and conclude that asymmetry is pervasive and large, however, we must be concerned that the findings are sensitive to model specification.

NOTES

While retaining responsibility for any errors or omissions in the analysis, the authors thank Jim Dolmas, Fred Joutz, Evan Koenig, Jayeong Koo, Don Norman, and Marci Rossell for helpful comments on earlier drafts of the paper, and Carrie Kelleher and Dong Fu for able research assistance.

1 We conceptualize the relationship between upstream and downstream prices as a markup model but conduct our estimation in levels and natural logs. Although the results are substantially similar for both specifications, we report these results for natural logs because that specification is scale invariant.

2 Causality tests conducted with forms of the model that allowed for asymmetry yielded substantially similar results.

3 We use a Choleski decomposition that decomposes the residuals $\mu_{1,t}$ and $\mu_{2,t}$ into two sets of impulses that are orthogonal to each other. This permits the covariance between the residuals to be taken into account. The Choleski decomposition imposes a recursive structure on the system of residuals in which the ordering of the residuals associated with each dependent variable is specified. If the covariance between the residuals is sufficiently high, the ordering can affect the results. We found that changing the ordering had little effect on the results, except the pairing of spot crude oil with spot gasoline.

4 The inclusion of a contemporaneous upstream price term raises a concern about the possibility of simultaneous equation bias. The upstream origin of the shocks mitigates much of this concern, and BCG found that failure to instrument the variable has no appreciable effect on the results.

5 Statistical tests indicate that the seasonal dummies are significant in all regressions and the time-trend variable is significant in some regressions. Robustness checks indicate that the seasonal dummies and the time-trend variable have little effect on the results.

6 The presence of the dummies, U_{t-i} and D_{t-i}, prevents us from rewriting the asymmetric differenced terms as levels terms without placing restrictions on the resulting coefficients. See Equation 7 below.

7 The differences between Equations 4 and 6 are best seen in Equation 7 below.

8 Not all figures are presented here. The remaining figures are available from the authors.

9 Because the models are nonlinear, some care must be taken in computing these responses. For all the responses, we consider a one-unit change in the upstream price, given that the upstream price is initially equal to its sample mean. Because lagged values of the downstream price enter the model, downstream prices are set equal to the steady value implied by the model when the upstream price is equal to its sample mean. The confidence bands are calculated by Monte Carlo Integration. For each replication, we randomly draw the model parameters, $\tilde{\beta}$, from its posterior distribution, which is assumed to be $N(\hat{\beta}, V(\hat{\beta}))$, where $\hat{\beta}$ and $V(\hat{\beta})$ are the estimated parameters and their variance-covariance matrix, respectively. For a given realization of $\tilde{\beta}$, we then calculate the responses of the downstream price to an increase and a decrease in the upstream price. This is repeated 1,000 times to form the two-standard-deviation confidence band.

REFERENCES

Bacon, Robert W. (1991), "Rockets and Feathers: The Asymmetric Speed of Adjustment of U.K. Retail Gasoline Prices to Cost Changes," *Energy Economics* 13 (July): 211–18.

Borenstein, Severin (1991), "Selling Costs and Switching Costs: Explaining Retail Gasoline Margins," *Rand Journal of Economics* 22 (Autumn): 354–69.

Borenstein, Severin, A. Colin Cameron, and Richard Gilbert (1997), "Do Gasoline Prices Respond Asymmetrically to Crude Oil Prices?" *Quarterly Journal of Economics* 112 (February): 305–39.

Deltas, George (1997), "Retail Gasoline Price Response Asymmetries to Wholesale Price Shocks" (Paper presented at the Western Economic Association Meeting, Seattle, July 9–13).

French, Mark (1991), "Asymmetry in Gasoline Price

Changes" (Washington, D.C.: Board of Governors of the Federal Reserve System, August, Draft paper).

Karrenbrock, Jeffrey D. (1991), "The Behavior of Retail Gasoline Prices: Symmetric or Not?" Federal Reserve Bank of St. Louis *Review*, July/August, 19 – 29.

Mohanty, Samarendu, E. Peterson, F. Wesley, and Nancy Cottrell Kruse (1995), "Price Asymmetry in the International Wheat Market," *Canadian Journal of Agricultural Economics* 43 (November): 355 – 66.

Neumark, David, and Steven A. Sharpe (1992), "Market Structure and the Nature of Price Rigidty: Evidence from the Market for Consumer Deposits," *Quarterly Journal of Economics* 107 (May): 657– 80.

Norman, Donald A., and David Shin (1991), "Price Adjustment in Gasoline and Heating Oil Markets," American Petroleum Institute Research Study no. 060 (Washington, D.C., August).

Scherer, F. M. (1980), *Industrial Market Structure and Economic Performance*, 2nd ed. (Chicago: Rand McNally).

Shin, David (1992), "Do Product Prices Respond Symmetrically to Changes in Crude Oil Prices," American Petroleum Institute Research Study no. 068 (Washington, D.C., December).

Tirole, Jean (1990), *The Theory of Industrial Organization* (Cambridge, Mass.: MIT Press).

U.S. General Accounting Office (1993), "Energy Security and Policy: Analysis of the Pricing of Crude Oil and Petroleum Products," GAO/RCED-91-17 (Washington, D.C.: Government Printing Office, March).

Has NAFTA Changed North American Trade?

David M. Gould
Senior Economist and Policy Advisor
Federal Reserve Bank of Dallas

Typically, the debate over NAFTA has focused on jobs. However, to really understand NAFTA's effects on employment or living standards, it is important to first answer the more fundamental question of what effect it has had on trade.

The North American Free Trade Agreement has been one of the most hotly debated trade accords in recent history. NAFTA's critics regard the expansion of free trade to a developing country like Mexico as a dangerous precedent. They envision U.S. jobs lost in a flood of goods from a country with an average wage one-fifth that of the United States. Others see NAFTA as a boon to U.S. employment and living standards through greater trade and investment opportunities.[1]

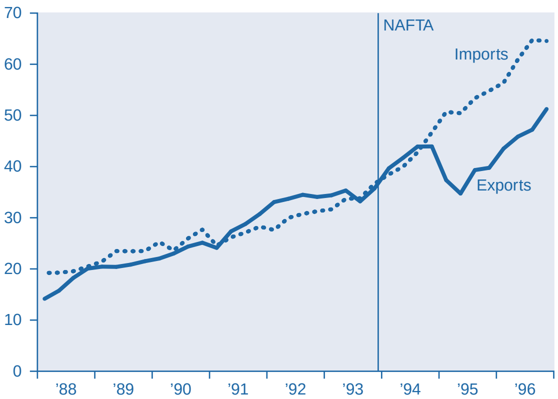
These opposing expectations for NAFTA have been largely matters of speculation and based on assessments of other trade accords. But now that NAFTA has been in operation for more than three years, the question is not what the trade accord is likely to do but what it has done.

Typically, the debate over NAFTA has focused on jobs. However, to really understand NAFTA's effects on employment or living standards, it is important to first answer the more fundamental question of what effect it has had on trade. Changes in trade patterns caused by a lowering of trade barriers are ultimately the mechanism by which jobs and living standards are influenced. This article examines how NAFTA, since its inception, has affected trade between the United States, Canada, and Mexico, holding constant other important factors that affect trade. Without controlling for these other factors, the effects of NAFTA are difficult to discern, which can lead to wrong or conflicting conclusions about the accord's effects on trade.

In the three years since NAFTA's implementation, both its supporters and opponents have used changes in the pattern of trade flows to justify their positions. Supporters have argued that during 1994, the year NAFTA took effect, U.S. trade with Mexico grew nearly 10 percent faster than the average of the previous five years (*Figure 1*). Opponents claim that any expansion in trade in NAFTA's first year was quickly reversed when expectations about its benefits fell to earth with the 1995 peso crisis. During 1995, U.S. imports from Mexico grew nearly 25 percent, but exports dropped 11 percent.

Since Mexico began to recover from its deep recession in late 1995, U.S. exports to Mexico have resumed rapid growth, but claiming the success or failure of NAFTA based on a superficial examination of the ups and downs of trade flows can be a mistake. The acccord's effects may be much more or less than a simple glance at these flows would suggest because NAFTA does not exist in an economic or policy vacuum.

Figure 1
U.S. Trade with Mexico



SOURCE: International Monetary Fund, Direction of Trade Statistics.

Worldwide economic changes that likely influenced bilateral trade within North America were already under way when NAFTA took effect. For example, U.S. real gross domestic product increased 3.5 percent in 1994, influencing the United States' supply and demand for imports and exports worldwide. As Figure 2 shows, U.S. trade with the world, excluding Mexico and Canada, grew faster in 1994 than in the previous six years. Likewise, Mexican real GDP increased 5.2 percent, and the real value of the peso was quite high in 1994, both factors that would have boosted U.S. exports to Mexico. As a result, it is unlikely that NAFTA and its lower trade barriers were the only influence on bilateral trade flows. To isolate the effects of NAFTA, one must account for the effects of changes in income, exchange rates, and trade with other countries; only then can NAFTA's impact on trade be ascertained.

In the first section of this article, I discuss previous analyses of NAFTA and compare them with my methodology. I next specify and estimate a model of the accord's effects on North American trade flows, and I assess how much trade has been influenced by NAFTA. I conclude by evaluating NAFTA's success.

DETERMINING NAFTA'S EFFECTS

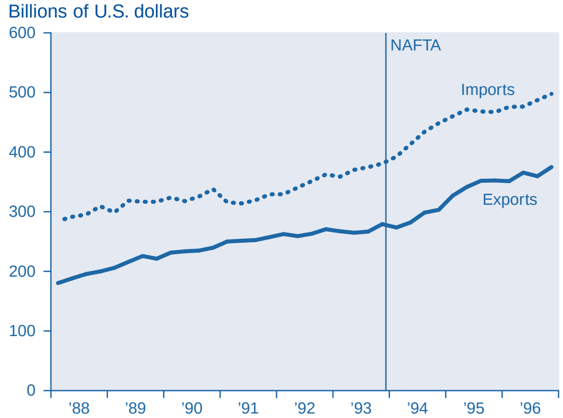
Because most studies of how NAFTA affects jobs, trade, and incomes in North America were completed just before, or shortly after, the treaty's implementation, the majority of them are forward-looking. This contrasts with the backward-looking approach of more recent studies, including this one.[2] However, any assessment of NAFTA's effects, past or future, must use an economic model to judge how trade interacts with the larger economy.

The techniques used to determine NAFTA's future impact fall into two broad categories. First is the computable general equilibrium (CGE) modeling technique, in which analysts model a simplified economy and simulate what would happen to it if tariffs and nontariff barriers fall according to the NAFTA schedule.[3] Typically, CGE models are static—that is, the structure of the economy cannot vary over time in response to changing trade patterns. More recently, however, a model by Kouparitsas (1997) allows for changes in capital investment and its reallocation across economic sectors and countries over time. Allowing capital flows to change over time in response to NAFTA generates a much larger benefit to freer trade.

Although these two types of studies make different assumptions about the economy's structure and the degree of competition in various sectors, both find an increase in income and trade under NAFTA. There is also substantial agreement about how NAFTA will affect Canada, Mexico, and the United States. In general, because Mexico's is the smallest economy, it reaps the largest benefit relative to its GDP (Kehoe and Kehoe 1994). The United States benefits modestly, and Canada does not have much benefit beyond that resulting from its 1989 free trade agreement with the United States.

Other studies use a partial equilibrium analysis to examine the effects of NAFTA. That is, they focus on particular sectors and assume

Figure 2
U.S. Trade with the World, Less Mexico and Canada



SOURCE: International Monetary Fund, Direction of Trade Statistics.

that the rest of the economy is unaffected. Determining NAFTA's trade effects in this context usually entails multiplying reduced-form price elasticities (a measure of how price changes translate into quantity changes) by the expected changes in tariff and tariff-equivalent trade barriers to determine how trade in various sectors is likely to change. In 1993, the U.S. International Trade Commission (USITC) completed a comprehensive study of NAFTA that used reduced-form price elasticities to determine how much NAFTA would affect trade in various U.S. economic sectors.[4] Hufbauer and Schott (1993) discuss NAFTA in a broad context and, based on their expectations of how trade would change under NAFTA, estimate how jobs in various sectors may be affected. They estimate changes in the number of jobs using Department of Commerce data on jobs supported directly and indirectly by exports to Mexico in 1990. They find that NAFTA and Mexico's economic reforms would increase the net number of U.S. jobs by about 170,000 in 1995 (Hufbauer and Schott 1993, 14).

In general, these partial equilibrium studies estimate that the United States will increase its exports of high-tech goods, grains and oilseed, and mechanical parts used in Mexican assembly plants. Increases are projected for U.S. imports of automobiles, apparel, glassware, household appliances, and certain horticultural products.

NAFTA SINCE ITS IMPLEMENTATION

Unlike the studies cited above, which have sought to predict how NAFTA will affect trade, this article's intent is to measure how NAFTA has already affected it. To assess the effects of NAFTA in its first three years, a model of bilateral trade flows in North America is empirically estimated with pre- and post-NAFTA data. The analysis is based on a widely used model of bilateral trade flows that includes incomes, prices, and exchange rates (see the box entitled "The Gravity Model of Bilateral Trade"). Once the fundamental determinants of trade flows are accounted for, any extraordinary flows that have occurred since NAFTA's inception are attributed to the free trade agreement.[5]

This type of analysis has at least two benefits. First, the analysis can estimate trade flows utilizing data since NAFTA's implementation. Second, it can capture most of the potentially important aspects of NAFTA, including tariff and nontariff barrier reductions, as well as changes in administrative rules, regulations, and expectations about the sustainability of free trade that cannot be easily quantified. Changing expectations for the sustainability of free trade under NAFTA are potentially among the most important aspects of the agreement. If NAFTA did not create a credible commitment to free trade, new investment would not flow into export industries to take advantage of reduced trade barriers. Without a credible free trade agreement, the benefits of the trade accord would be much lower. Expectations for a more stable and open trading environment affect trade by providing the incentive for firms to make long-term capital commitments. These expectations cannot be easily accounted for using the methodology cited previously because it is difficult to translate them into price changes.

Of course, there are potential pitfalls to this methodology as well. In particular, the model is unlikely to control for all the factors important to bilateral trade flows. Factors may be attributed to NAFTA that should not be. For example, when NAFTA took effect on January 1, 1994, political unrest was developing in Mexico, resulting in the armed Zapatista movement and two political assassinations later that year. Inasmuch as the uncertainty generated by this unrest reduced trade, it would reduce the estimated trade effects of NAFTA. Even though the events are independent of NAFTA, they would be indistinguishable in the context of the empirical model. If political uncertainty could be measured and included in the model, this would not be a problem.

However, even if important economic or political events are excluded from the model, they may not bias the estimated NAFTA effects if they and the accord's implementation were not simultaneous. Factors such as the peso crisis are unlikely to bias the analysis because exchange rates and incomes (the two factors most affected by the crisis) are included in the model. In other words, a majority of the peso crisis' effects on trade are likely to be taken into account.[6] Still, because NAFTA was implemented in a rather tumultuous period in Mexico that resulted in reduced trade, the estimated effects of NAFTA may be biased downward.

ESTIMATING THE EFFECTS OF NAFTA

To assess the effects of NAFTA since its implementation, the following benchmark model of Canadian, Mexican, and U.S. bilateral trade flows is estimated using quarterly data

The Gravity Model of Bilateral Trade

The empirical bilateral trade model in this article is based on the gravity equation, which derives its name from its resemblance to Newton's law of gravity. The model was originally formulated from ad hoc assumptions but had its intuitive appeal in describing trade flows as increasing as the economic "distance" between two countries shrinks. It describes bilateral trade flows between two countries as a function of their incomes, populations, the physical distance between them, and trade barriers. If countries with similar incomes have similar preferences for goods but produce different types of products, they are likely to trade more with each other than with other countries. Trade is also likely to increase the closer the countries are and the lower the trade barriers between them. The gravity model has been used to describe many different types of flows, such as immigration, shopping patterns, and car traffic, as well as interregional trade. It has been used extensively in international trade applications because it provides an empirically tractable framework.

The ad hoc assumptions behind the gravity equation have been replaced by microeconomic foundations. Anderson (1979), Helpman and Krugman (1985), Bergstrand (1985), and Bikker (1987) have developed variants of the gravity model based on utility and profit maximization. The empirical model this article uses is based on Bergstrand's theoretical foundation for the gravity equation, which is based on the assumption that producers maximize profits subject to a constant elasticity of transformation (CET) technology, and consumers maximize a constant elasticity of substitution (CES) utility function subject to a budget constraint (Bergstrand 1985). Assuming that individual bilateral trade flows are small relative to total trade, the equation for bilateral trade is

$$
\begin{aligned}
\text{(B.1)} \qquad PX_{ij} = {} & Y_i^{(\sigma-1)/(\gamma+\sigma)} Y_j^{(\gamma+1)/(\gamma+\sigma)} T_{ij}^{-\sigma(\gamma+1)/(\gamma+\sigma)} C_{ij}^{-\sigma(\gamma+1)/(\gamma+\sigma)} E_{ij}^{\sigma(\gamma+1)/(\gamma+\sigma)} \\
& * \left(\sum_{k=1,k\neq i}^{N} P_{ik}^{1+\gamma} \right)^{-(\sigma-1)/(\gamma-\eta)/(1+\gamma)(\gamma+\sigma)} \\
& * \left(\sum_{k=1,k\neq j}^{N} P_{kj}^{*1-\sigma} \right)^{(\gamma+1)(\sigma-\mu)/(1-\sigma)(\gamma+\sigma)} \\
& * \left[\left(\sum_{k=1,k\neq i}^{N} P_{ik}^{1+\gamma} \right)^{(1+\eta)/(1+\gamma)} + P_{ii}^{1+\eta} \right]^{-(\sigma-1)/(\gamma+\sigma)} \\
& * \left[\left(\sum_{k=1,k\neq j}^{N} P_{kj}^{*1-\sigma} \right)^{(1-\mu)/(1-\sigma)} + P_{jj}^{1-\mu} \right]^{-(\gamma+1)/(\gamma+\sigma)},
\end{aligned}
$$

where PX_{ij} is the value of aggregate trade flows from country i to country j,
Y_i is the aggregate income of country i,
Y_j is the aggregate income of country j,
P_{ik} is the price received for country i's product in the k^{th} country,
$P^*_{kj} = P_{kj}T_{kj}/E_{kj}$ is the price paid for buying k's product in the j^{th} market,
T_{ij} is 1 plus the ad valorem tariff rate on i's product in the j^{th} market,
C_{ij} is a nontariff barrier on i's product in the j^{th} market,
E_{ij} is the exchange rate between country j's currency in terms of i's currency,
γ is the constant elasticity of transformation in the supply between different export goods ($0 \leq \gamma \leq \infty$),
η is the CET between the supply of exports and domestically produced goods ($0 \leq \eta \leq \infty$),
σ is the constant elasticity of substitution between the demand for different imported goods ($0 \leq \sigma \leq \infty$), and
μ is the CES between the demand for imported and domestic goods ($0 \leq \mu \leq \infty$).

As shown in Equation B.1, the value of aggregate trade flows from country i to country j depends on nine terms. In the order of their appearance in the equation, they are (1) the income of the exporting country, (2) the income of the importing country, (3) tariffs, (4) nontariff barriers, (5) the exchange rate, (6) an export price index for exports to all countries to which the exporting country exports, (7) an import price index for imports from all countries from which the importing country imports, (8) an index of domestic prices for the exporting country, and (9) an index of domestic prices for the importing country.

These nine terms can be sorted into three categories: (1) income in the exporting and importing countries, which reflects the potential demand and supply; (2) wedges between the export and import price of the traded goods due to tariffs and nontariff barriers; and (3) price terms reflecting the substitutability between this traded good and the others.

Equation B.1 serves as the basis for the empirical model describing trade flows between the NAFTA countries. Changes in tariff and nontariff barriers are proxied by a binary variable for NAFTA. Because country-specific data for the price terms are not available, the empirical analysis uses proxies. To account for the exchange rate, domestic prices, and the terms of trade between the bilateral trade partners, GDP price deflators and the bilateral real exchange rate between the partners are used. To account for the terms of trade with other trading partners, a multilateral real exchange rate with the rest of the world is used. Economic events, such as the beginning of Mexico's trade liberalization in 1985 and Canada's free trade agreement with the United States in 1989, are proxied by binary variables.

from 1980 though 1996. The empirical equations are based on the gravity model, which is derived from standard microeconomic foundations (Bergstrand 1985). All variables are seasonally adjusted quarterly data and are expressed in log first-differences (growth rates):[7]

$$(1)\quad M_t^{ij} = \alpha_0 + \alpha_1 M_{t-q}^{ij} + \alpha_2 Y_{t-q}^{i} + \alpha_3 Y_{t-q}^{j} + \alpha_4 E_{t-q}^{ij} + \alpha_5 E_{t-q}^{iw} + \alpha_6 P_{t-q}^{i} + \alpha_7 P_{t-q}^{j} + \alpha_8 D_t + \alpha_9 NAFTA_t + \epsilon_t;$$

$$(2)\quad X_t^{ij} = \beta_0 + \beta_1 X_{t-p}^{ij} + \beta_2 Y_{t-p}^{i} + \beta_3 Y_{t-p}^{j} + \beta_4 E_{t-p}^{ij} + \beta_5 E_{t-p}^{iw} + \beta_6 P_{t-p}^{i} + \beta_7 P_{t-p}^{j} + \beta_8 D_t + \beta_9 NAFTA_t + \mu_t.$$

The variables are defined as follows: M^{ij} is country i's imports from country j; i and j are either the United States, Canada, or Mexico; t refers to the date; p and q are the number of periods a variable is lagged; X^{ij} is country i's exports to country j; Y^i is real GDP of country i; Y^j is real GDP of country j; P^i is country i's GDP price deflator; P^j is country j's price deflator; E^{ij} is the real exchange rate between countries i and j; E^{iw} is the real exchange rate between country i and the world (excluding country j); and D is a binary variable that represents changes in trade regimes not associated with NAFTA. For trade with Mexico, D equals 1 beginning in 1985, the period in which Mexico begins liberalizing trade. For trade between Canada and the United States, D equals 1 beginning in 1989, representing the period of the U.S.–Canada free trade agreement. *NAFTA* is a binary variable representing the period in which the accord was implemented. The variable *NAFTA* equals 1 beginning the last quarter of 1993.[8] α and β are estimated coefficients, and ϵ and μ are error terms.

The variables in Equations 1 and 2 can be sorted into four categories: (1) lagged trade (M_{t-p}^{ij}, X_{t-q}^{ij}), which reflects the adjustment process of trade to a new equilibrium; (2) income in the exporting and importing countries (Y^j, Y^i), which reflects the potential demand and supply for the traded goods; (3) price and real exchange rate terms (P^i, P^j, E^{ij}, E^{iw}), reflecting the substitutability of nontraded and traded goods in the NAFTA countries and the rest of the world; and (4) one-time trade liberalization variables, reflecting changes in trade regimes and NAFTA (*D, NAFTA*).

The size and statistical significance of the coefficient on the *NAFTA* variable tell us the degree to which NAFTA affects bilateral trade flows in North America.[9] It should be noted that NAFTA is not scheduled to be fully implemented until 2009. Tariff rates in many sectors are to be reduced over a fifteen-year period (see the box entitled "What Has NAFTA Done?"). Consequently, these results should be seen as a preliminary look into NAFTA's effects on aggregate trade.

The estimated equations are in the appendix.[10] Overall, the equations explain the growth of trade relatively well.[11] However, the effects of NAFTA on trade flows (in size and statistical significance) vary a great deal between countries. Figures 3 through 8 show what the estimation results imply for actual exports and imports between the United States, Canada, and Mexico. The shaded bands on both sides of trade estimated without NAFTA represent a 90 percent confidence interval derived from the statistical error of the estimate.[12]

The United States and Mexico

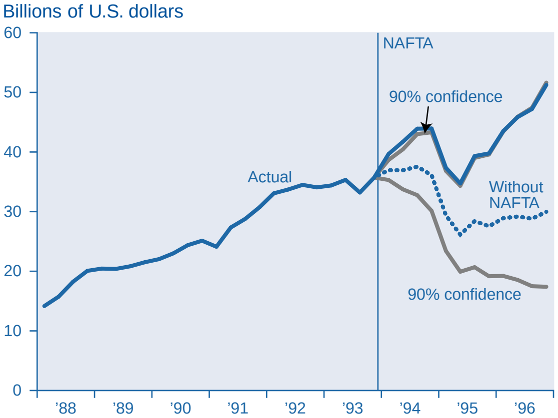
Figures 3 and 4 show NAFTA's estimated effects on bilateral trade flows between the United States and Mexico. As the dotted line in Figure 3 indicates, U.S. exports are estimated to have grown faster than they would have had there not been a trade agreement. On average, U.S. export growth is about 16.3 percentage points higher per year with NAFTA. While the increase in growth is not extraordinary, the cumulative effect is about $21.3 billion more in exports than what would have occurred without NAFTA. The statistical significance of this effect is high, as shown by the 90 percent confidence interval lines that exclude the observed data on U.S. exports to Mexico.

For U.S. imports, as shown in Figure 4, the boost from NAFTA is also relatively high. On average, import growth is about 16.2 percentage points higher per year with NAFTA. Since NAFTA became law, the cumulate impact amounts to about $20.5 billion in additional imports because of the agreement. However, NAFTA's statistical significance for U.S. imports from Mexico is only marginal. The 90 percent confidence interval lines show that we cannot exclude the possibility that trade without NAFTA would have been different from trade with NAFTA.

The United States and Canada

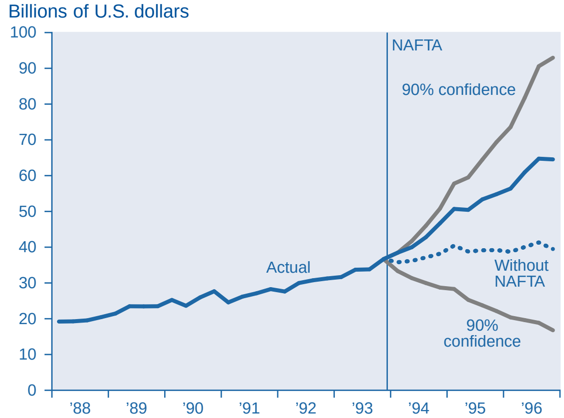
Figures 5 and 6 show the estimated effects of NAFTA on bilateral trade flows between the United States and Canada. As both figures show, trade between the two countries has not been affected much by NAFTA. This is not surprising, given that a free trade agreement with Canada was negotiated in 1989 and NAFTA did not alter

Figure 3
U.S. Exports to Mexico



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

Figure 4
U.S. Imports from Mexico



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

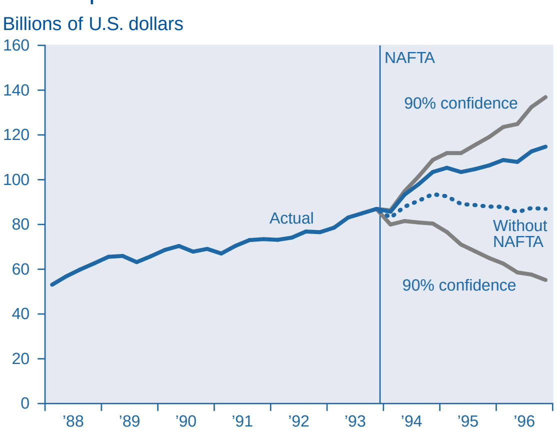
that accord significantly. However, even the 1989 agreement, as measured by the binary variable D^{ij} in the U.S.–Canada trade equations, does not seem to play a strong role in determining trade flows. Although trade may have been liberalized in some sectors, aggregate trade does not seem to be influenced much.[13] This may be because trade between the countries has been generally open for some time. Figures 5 and 6 show an 8.6 percent average annual increase in U.S. exports to Canada and a 3.9 percent increase in U.S. imports from Canada due to NAFTA. The NAFTA effect on both exports and imports is statistically insignificant.

Canada and Mexico

The effects of NAFTA on exports and imports between Canada and Mexico appear in Figures 7 and 8. As the figures show, the estimated effects are very imprecise, with a wide, 90 percent confidence band. One possible reason for the difficulty in measuring the effects of NAFTA in these equations is that trade between Canada and Mexico is a very small share of each country's total and is subject to much more unexplained volatility than is trade with the United States.[14] In these equations, the NAFTA trade effects are estimated to be negative, which raises the possibility that NAFTA may have diverted Canadian–Mexican trade toward the United States or other countries. But because the effects are so imprecise, the possibility that the effects are zero or even positive cannot be excluded.

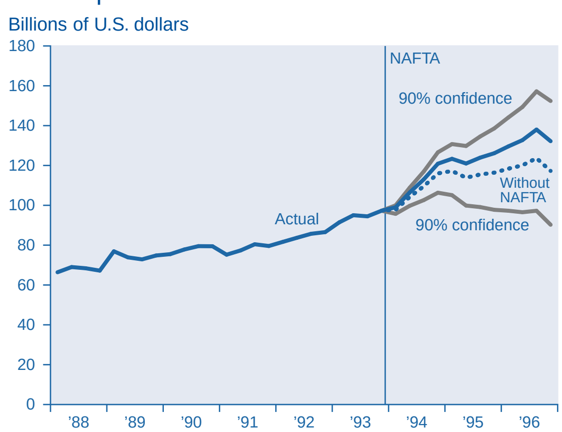
In summary, it is important to remember that there is a wide statistical margin of error for most of the estimated NAFTA trade effects, so they should be viewed in relative rather than abso-

Figure 5
U.S. Exports to Canada



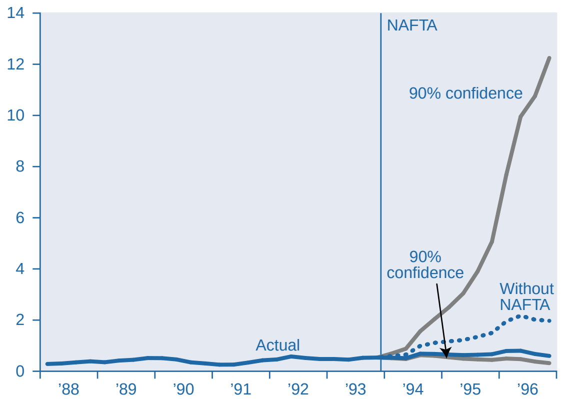
SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

Figure 6
U.S. Imports from Canada



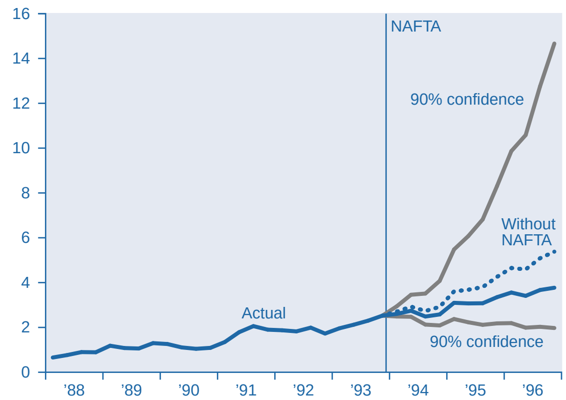
SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

Figure 7
Canadian Exports to Mexico



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

Figure 8
Canadian Imports from Mexico



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

lute terms. While it is likely that NAFTA affected U.S.–Mexican trade, it is unlikely that it affected U.S.–Canadian or Canadian–Mexican trade.

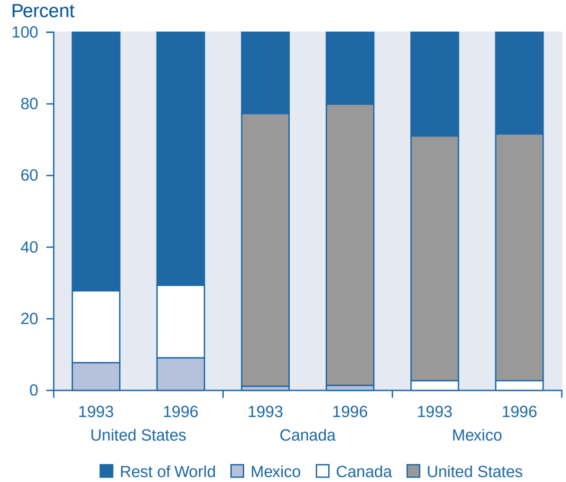
TRADE DIVERSION VERSUS TRADE CREATION

To judge NAFTA's effects on the economies of Canada, Mexico, and the United States, it is also important to consider North American trade flows in the context of trade with the rest of the world. In other words, did NAFTA create new trade opportunities within North America, or did it simply divert trade from countries outside NAFTA? If the increased trade caused by NAFTA was simply a shuffling of trade from other, more efficient trading partners, then NAFTA's benefit would shrink. Although a detailed examination of this issue is not within the scope of this article, a glance at how the distribution of trade flows has changed since NAFTA can tell us whether the accord may be associated with trade diversion.

Figure 9 shows how the distribution of trade flows between the NAFTA countries and the rest of the world changed from 1993, the year before NAFTA began, to 1996, three years after NAFTA started. As the figure shows, trade within North America has increased relative to trade with the rest of the world, but the increase is slight. The share of U.S. trade with Canada and Mexico increased from 27.8 to 29.4 percent between 1993 and 1996, with most of that increase attributed to greater U.S. trade with Mexico. Canadian trade with the rest of North America also increased, from a share of 77.3 to 80.4 percent. Mexico's trade share with North America changed very little, from 71 to 71.6 percent.[15]

The share of total trade between the NAFTA countries slightly increased, suggesting that if there was trade diversion, it was small. But to determine the extent of trade diversion in North America, it is also important to consider whether increased trade between NAFTA countries came at the expense of trade with the rest of the world. In other words, did NAFTA shift trade away from countries outside of NAFTA, or did NAFTA simply increase trade within North America at a faster rate than trade increased with the rest of the world? Figure 10 shows that it was the latter; that is, trade with countries outside North America also grew after NAFTA's implementation. The share of total trade between North American countries increased

Figure 9
Trade Shares of NAFTA Countries, Before and After the Treaty's Inception



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

What Has NAFTA Done?

On January 1,1994, NAFTA substantially reduced trade barriers across a market with more than 380 million people and a combined gross domestic product of roughly $7.6 trillion. Although trade barriers have already been lowered significantly, NAFTA will not be completely phased in until 2009. Most of the declines in tariffs have been on the Mexican side because Mexico started with higher tariffs than the United States or Canada. In 1992, Mexican tariffs on imports from the United States averaged about 10 percent when weighted by the value imported; at the same time, U.S. tariffs on imports from Mexico averaged about 4 percent. Because Canada and the United States negotiated a separate free trade agreement in 1989, NAFTA affected trade between the two countries very little.

NAFTA substantially reduces, but does not eliminate, nontariff trade barriers, such as import quotas, sanitary regulations, and licensing requirements. Canada and the United States traditionally have had few restrictions on capital flows, whereas Mexican laws prohibited private ownership in the petroleum industry and parts of the petrochemical industry, restricted foreign investment in the financial and insurance sectors, and institutionalized communal ownership of agricultural lands. The petroleum industry is still off-limits to foreign investment, although parts of the petrochemical industry are set to be privatized. Many laws against foreign investment in the financial and insurance sectors have been eliminated or substantially reduced. Although NAFTA set a schedule for liberalizing the banking sector, the 1995 peso crisis helped generate the political momentum to speed the opening of this sector by decreasing the restrictions on foreign ownership of existing banks.

NAFTA has not been without glitches, but the problems are probably fewer than what they would have been without the agreement. The opening of trucking between the United States and Mexico has been delayed, and tariffs have increased for some products. The United States levied additional tariffs on Mexican straw brooms, and in response Mexico levied tariffs on U.S. alcoholic beverages, flat glass, notebooks, and some types of wood furniture. There have also been disputes over agricultural products, such as avocados and tomatoes. Although these disputes are troublesome, their effect on overall trade has been small. Moreover, NAFTA may have limited a protectionist response to the 1995 peso crisis. Unlike Mexico's 1982 crisis, when the Mexican government raised tariffs dramatically in the hope of generating a trade surplus to boost foreign reserves, during the 1995 crisis no such political response occurred.

Here are some specifics by sector on how NAFTA has reduced trade barriers.[1]

Automobiles

NAFTA immediately decreased Mexican tariffs on automobiles from 20 to 10 percent in 1994 and is set to drop them to zero by 2004. Tariffs on most auto parts will be eliminated by 1999. The agreement includes rules of origin specifying that to qualify for preferential tariff treatment, vehicles must have 62.5 percent North American content, which is an increase over the 50 percent provision in the U.S.–Canadian free trade agreement.

By 2004 NAFTA eliminates requirements that automakers supplying the Mexican market produce the cars in Mexico and buy Mexican parts. It has already eliminated mandatory export quotas on foreign-owned auto manufacturing facilities in Mexico, and by 1999 it eliminates the Mexican restriction on bus and truck imports.

Textiles and Apparel

NAFTA immediately eliminated trade barriers on more than 20 percent of Mexican–U.S. trade in textiles and apparel. Over six years it eliminates barriers on another 60 percent. The accord's rules of origin require that, to receive NAFTA tariff preferences, apparel be manufactured in North America from the yarn-spinning state forward.

Agriculture

NAFTA immediately reduced tariffs to zero for half of U.S. agricultural exports to Mexico. The other half of agricultural goods tariffs are to be eliminated by 2009. NAFTA immediately eliminated Mexico's licensing requirements for grains, dairy, and poultry.

Financial Services

NAFTA immediately reduced, and will eliminate by 2000, Mexico's restrictions on Canadian and U.S. ownership and provision of commercial banking, insurance, securities trading, and other financial services. Under NAFTA, Canadian and U.S. financial firms are allowed to establish wholly owned subsidiaries in Mexico and to engage in the same range of activities as similar Mexican firms.

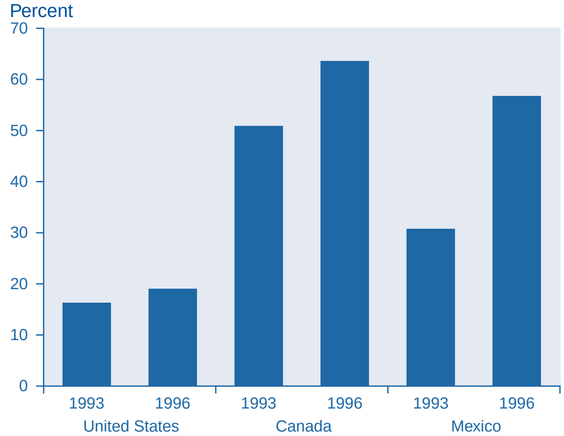
[1] Much of the following is described in Kehoe and Kehoe (1994).

because trade within North America grew faster than did trade with countries outside of North America. Consequently, although trade diversion is a possibility, it is unlikely to be a large problem. Moreover, because trade under NAFTA was liberalized between countries with very different comparative advantages, it is unlikely that it caused a shift from optimal trading patterns.

HAS NAFTA BEEN A SUCCESS?

Certainly, NAFTA is not the solution to all the economic problems that ail North America, but it is not the disaster that critics claimed it would be. NAFTA is foremost a free trade agreement, and as such its benefits derive from a shift in resources to industries that reflect a nation's comparative advantage and away from indus-

Figure 10
Trade with the World as a Share of GDP
(Excluding trade between NAFTA countries)



SOURCES: International Monetary Fund, Direction of Trade Statistics and International Financial Statistics.

tries that do not. It is important to understand that this shift implies that the benefits come from both increased imports and exports. Accordingly, the best way to judge a free trade agreement is by whether it increases imports and exports, and not by whether it increases exports and decreases imports. By this criterion, NAFTA has been a success for the United States and Mexico. As expected, NAFTA has meant little for the Canadian economy.

After accounting for the effects of economic variables important to bilateral trade flows—such as income, exchange rates, and prices—NAFTA is found to have a significant positive effect on trade flows between the United States and Mexico. NAFTA is not found to have a significant impact on trade between the United States and Canada or Canada and Mexico. These findings are not surprising, given that the United States negotiated a free trade agreement with Canada five years before the implementation of NAFTA and that most of the trade liberalized under NAFTA is between the United States and Mexico.

Although this empirical analysis controls for economic shocks that would affect trade through changes in incomes and exchange rates, such as the 1995 peso crisis, it cannot control for all external shocks, nor can it capture all aspects of NAFTA's influence on trade. Perhaps the largest omission from the analysis is trade barriers that were not erected because of the free trade agreement but would have been without it. This issue was particularly relevant during the 1995 peso crisis. Unlike previous periods of economic turmoil in Mexico, trade was relatively unimpeded during the peso crisis. NAFTA, by enhancing the economic ties between the North American countries, may have limited a protectionist response to the peso crisis and helped facilitate a return of foreign investment and economic growth to Mexico.

Has NAFTA destroyed U.S. jobs? Clearly, NAFTA has neither spelled the death of the U.S. workforce, nor has it generated a dramatic increase in the number of U.S. jobs. What dominates the employment picture in any year are movements in a country's own business cycle, not trade. U.S. income grew fairly smoothly between 1994 and 1996; as a result, U.S. employment grew by 3.6 million. In contrast, Mexico experienced a currency crisis and deep recession in 1995; its employment fell but is now recovering with the economy. Ultimately, freer trade does not determine the number of jobs available in a country, but it does determine the types of jobs available. In the three years since NAFTA's implementation, there has been a clear trend toward increased trade in North America and higher productivity in the United States. How much of that greater productivity is due to NAFTA is unknown. As time passes, and more economic data become available, cyclical factors and economic shocks will fall to the background and a clearer picture of NAFTA's effects on the economy will emerge.

NOTES

I thank Baoyuan Wang for excellent research assistance, and Evan Koenig, Bill Gruben, and Lori Taylor for their comments and suggestions. Any remaining errors are my own.

1 The early controversy can perhaps be best summarized by quotes from Ross Perot and President Bill Clinton during the NAFTA debate in 1993: "NAFTA will pit American and Mexican workers in a race to the bottom. In this race, millions of Americans will lose their jobs" (Perot 1993, *i*); "I believe the Nafta will create 200,000 American jobs in the first two years of its effect" (Clinton 1993).

2 See, for example, Gould (1996), Weintraub (1997), and USITC (1997).

3 For an excellent survey of general equilibrium models applied to NAFTA, see Kehoe and Kehoe (1994).

4 In creating the reduced-form price elasticities, the USITC study assumes that foreign and domestic goods are imperfect substitutes for each other. In other words, the goods have separate markets in which equilibrium prices and quantities are established. See USITC (1993) for a description of this methodology.

5 The methodology used here is an extension of the work done by Gould (1996). Other recent studies have used a similar methodology to assess the trade and sectoral effects of NAFTA. See USITC (1997).

6 See Gould (1996) for an assessment of how the peso crisis affected trade independent of NAFTA. See Neely (1996) for a discussion of why NAFTA did not cause Mexico's peso crisis.

7 Log first-differences, as opposed to a simple log-linear relationship, were used because tests on the dependent and many of the independent variables could not reject the hypothesis of nonstationarity. Consequently, the equations estimate the growth of exports and imports.

8 Because trade growth equations are estimated, the effects of NAFTA are assumed to influence the growth of trade. However, according to traditional long-run models of trade, lower tariffs only influence the level, not the growth, of trade. Because trade is unlikely to jump to a new, higher level instantaneously, the growth of trade is likely to be affected in the transition to a

new, higher level. This is especially true for the short period that NAFTA has been observed and because NAFTA is being phased in over fifteen years.

9 NAFTA may also affect bilateral trade flows indirectly through income and prices. Although these indirect effects are likely to be important over the long run, over the short run these effects are probably small. Because of this, these secondary effects are ignored in the estimation.

10 The equations were estimated with ordinary least squares and the errors terms checked to see if they follow a white-noise pattern. The lag structure of the equations was determined according to the Akaike information criterion. To determine how trade would have grown without NAFTA, the estimated *NAFTA* effect was excluded from the estimated exports and imports equations, and trade flows were calculated with the actual data for the independent variables. To provide the best estimates, the error term (which reflects the degree to which the equation does not match the data) was included in the calculation. Data sources are given in the appendix.

11 The adjusted R^2 on the equations varies from 0.67 in the U.S.–Mexico export equation to 0.18 in the Canada–Mexico import equation. Most of the equations have an R^2 between 0.30 to 0.40, which is not uncommon for similar growth equations. The adjusted R^2 measures the proportion of the variation in the left-side dependent variable that is explained by the right-side dependent variables, adjusting for the number of variables in the equation.

12 The confidence interval shows the degree of certainty we can have in the estimated effects. If the confidence interval around the estimated effects of trade without NAFTA excludes the actual observed trade under NAFTA, we can say with 90 percent certainty that trade with NAFTA is different from trade without it. If the 90 percent confidence interval includes the observed trade under NAFTA, we can say that there is less than a 90 percent certainty that trade is different with NAFTA than without it.

13 An inherent problem in studying aggregate exports and imports is that the analysis cannot explain changes in sector-specific trade flows caused by NAFTA. For example, imports in one industry may expand, while imports in another industry may contract. In aggregate, however, imports overall would appear to remain stable. An attempt was made to study sector-specific trade data, but because equivalent sector-specific price information across countries does not exist, the empirical results were poor.

14 This is indicated by the relatively low adjusted R^2 of the Canada–Mexico trade equations.

15 A bilateral trade intensity index, defined as the share of country *j*'s trade in country *i*'s world trade relative to the share of country *i*'s world trade in total world trade [$I_{ij} = (T_{ij}/T_{iw})/(T_{jw}/T_w)$], also shows a slight increase among NAFTA partners since 1993. See Yeats (1997) for a discussion of this index applied to Mercosur's trade.

REFERENCES

Anderson, James (1979), "A Theoretical Foundation for the Gravity Equation," *American Economic Review* 69 (March): 106.

Bergstrand, Jeffrey (1985), "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence," *Review of Economics and Statistics* 67 (August): 474.

Bikker, Jacob A. (1987), "An International Trade Flow Model with Substitution: An Extension of the Gravity Model," *Kyklos* 40 (3): 315–37.

Clinton, Bill (1993), quoted in Bob Davis, "Two Years Later, the Promises Used to Sell Nafta Haven't Come True, But Its Foes Were Wrong, Too," *Wall Street Journal*, October 26, 1995, A24.

Gould, David M. (1996), "Distinguishing NAFTA from the Peso Crisis," Federal Reserve Bank of Dallas *Southwest Economy*, September/October, 6–10.

Helpman, Elhanan, and Paul Krugman (1985), *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy* (Cambridge, Mass.: MIT Press).

Hufbauer, Gary C., and Jeffrey J. Schott (1993), *NAFTA: An Assessment* (Washington, D.C.: Institute for International Economics).

Kehoe, Patrick J., and Timothy J. Kehoe (1994), "Capturing NAFTA's Impact with Applied General Equilibrium Models," Federal Reserve Bank of Minneapolis *Quarterly Review*, Spring, 17–34.

Kouparitsas, Michael A. (1997), "A Dynamic Macroeconomic Analysis of NAFTA," *Economic Perspectives* 21 (January/February): 14–35.

Neely, Christopher J. (1996),"The Giant Sucking Sound: Did NAFTA Devour the Mexican Peso?" Federal Reserve Bank of St. Louis *Review,* July/August, 33–47.

Perot, Ross (1993), *Save Your Job, Save Our Country: Why NAFTA Must Be Stopped—Now!* (New York: Hyperion).

U.S. International Trade Commission (USITC 1997), "The Impact of the North American Free Trade Agreement on the U.S. Economy and Industries: A Three-Year Review," USITC Investigation no. 332–81, June.

U.S. International Trade Commission (USITC 1993), *Potential Impact on the U.S. Economy and Selected Industries of the North American Free-Trade Agreement*, USITC Publication 2596 (Washington, D.C.: USITC, January).

Weintraub, Sidney (1997), *NAFTA at Three: A Progress Report* (Washington, D.C.: Center for Strategic and International Studies).

Yeats, Alexander (1997), "Does Mercosur's Trade Performance Raise Concerns About the Effects of Regional Trade Arrangements?" World Bank Policy Research Working Paper no. 1729 (Washington, D.C., February).

Appendix
Regression Results

Dependent variable	U.S.(i)–Mexico(j) Exports	U.S.(i)–Mexico(j) Imports	U.S.(i)–Canada(j) Exports	U.S.(i)–Canada(j) Imports	Canada(i)–Mexico(j) Exports	Canada(i)–Mexico(j) Imports
Constant term	.027	–.061	–.025	–.029	–.229	.169
	(.794)	(.713)	(.688)	(.078)	(.224)	(.042)
Lagged dependent variable	–.679	–.901	–.441	–.209	–.299	–.323
	(.074)	(.050)	(.659)	(.098)	(.341)	(.012)
Y^i	3.623	.289	3.953	2.793	9.474	2.560
	(.031)	(.209)	(.334)	(.135)	(.366)	(.543)
Y^j	.082	–3.046	–1.095	.104	3.677	.335
	(.050)	(.539)	(.394)	(.425)	(.255)	(.841)
E^{ij}	–2.147	–.610	–.171	.178	–.863	–.435
	(.000)	(.317)	(.536)	(.427)	(.053)	(.181)
E^{iw}	–2.971	–.061	–2.062	–.418	6.748	5.742
	(.047)	(.316)	(.430)	(.665)	(.497)	(.007)
P^i	.963	13.716	3.438	4.208	–17.100	–5.847
	(.616)	(.292)	(.106)	(.009)	(.384)	(.149)
P^j	–.533	–.191	.477	–.877	–1.185	–.890
	(.704)	(.608)	(.724)	(.578)	(.368)	(.099)
D^{ij}	–.029	.026	–.005	.008	–.008	–.003
	(.522)	(.722)	(.825)	(.550)	(.951)	(.994)
NAFTA	.073	.072	.031	.018	–.111	–.038
	(.015)	(.119)	(.191)	(.374)	(.240)	(.453)
Adjusted R^2	.67	.35	.36	.31	.27	.18
Equation F statistic (significance level)	.000	.030	.080	.002	.120	.050
LM test for autocorrelation (significance level)	.31	.53	.20	.25	.01	.24
Lag structure	3	5	5	1	4	1
Degrees of freedom	30	14	18	50	22	46

NOTE: Coefficients are the sum of the lagged terms. Significance level of F statistics (the null hypothesis that all lagged coefficients are equal to zero) are in parentheses.

Data Sources

Variable	Definition	Source
M	Seasonally adjusted value of merchandise imports, in millions of U.S. dollars	International Monetary Fund, Direction of Trade Statistics
X	Seasonally adjusted value of merchandise exports, in millions of U.S. dollars	International Monetary Fund, Direction of Trade Statistics
P	Seasonally adjusted GDP price deflator	International Monetary Fund, International Financial Statistics
Y	Seasonally adjusted real GDP	International Monetary Fund, International Financial Statistics
E	Seasonally adjusted real exchange rate	Trade-Weighted Value of the Dollar, Federal Reserve Bank of Dallas, and author's calculations
E^w	Seasonally adjusted real exchange rate with rest of the world	Trade-Weighted Value of the Dollar, Federal Reserve Bank of Dallas, and author's calculations

The Dynamic Impact Of Fundamental Tax Reform Part 1: The Basic Model

Evan F. Koenig
Senior Economist and Assistant Vice President
Federal Reserve Bank of Dallas

Gregory W. Huffman
Professor
Southern Methodist University
and
Research Associate
Federal Reserve Bank of Dallas

Adoption of a flat-rate consumption tax can be expected to have an immediate positive impact on saving and lead, in the long run, to higher levels of consumption, wages, and stock prices, and to lower interest rates. In the short run, however, real interest rates are likely to rise, and consumption and real stock prices are likely to fall.

Interest in fundamental tax reform has waned since the early months of the 1996 U.S. presidential campaign when it was the subject of intense political and media debate. In coming years, a resurgence of interest seems almost certain. After all, the Internal Revenue Service remains unpopular, the U.S. savings rate remains low, and pressure to efficiently raise substantial new tax revenues will grow once the baby boom generation reaches retirement age and federal entitlement spending begins to balloon. Now, while the rhetoric is still somewhat subdued, may be a good time to review the impact a major tax overhaul would have on the economy.

In this first of two articles on the economic impact of fundamental tax reform, we describe a framework useful for analyzing how the adoption of a flat-rate consumption tax would affect interest rates, the savings decision of a typical household, and the investment and hiring decisions of a typical firm.[1] We are less concerned with obtaining precise quantitative estimates of these effects than we are with establishing their direction and explaining the forces behind them. Moreover, our focus throughout is on the *macro*economic impact of tax reform. We will largely ignore the issues of who would be likely to gain most from reform and who might suffer losses. We take this approach partly because our analytical framework isn't well suited to addressing distributional questions, partly because the distributional impact of tax reform has already been adequately discussed elsewhere, and partly because windfall losses can often be reduced or eliminated through careful design of transition rules and the appropriate conduct of monetary policy.[2] In any event, a tax reform that succeeds in raising the economy's growth potential—even if only temporarily or only by a small amount—is likely to yield long-run net economic benefits to the vast majority of people. Our hope is that an improved understanding of the macroeconomic effects of tax reform will help readers keep potential short-term windfall losses in perspective.

Consistent with results obtained by others in complicated numerical simulation exercises, our analysis indicates that adoption of a flat-rate consumption tax can be expected to have an immediate positive impact on saving and lead, in the long run, to higher levels of consumption, wages, and stock prices, and to lower interest rates. In the short run, however, real interest rates are likely to rise, and consumption and real stock prices are likely to fall. These results are subject to important qualifications. First, our analysis ignores international capital flows. Such flows potentially exert a moderating influence on consumption and interest rate movements. However, different countries treat foreign-source income very differently for tax purposes, making it difficult to draw general conclusions.[3] Second, we ignore enforcement, avoidance, and administrative costs. Certainly, tax-reform advocates hope that these costs will fall. However, the magnitude of the cost savings will depend greatly on the specifics of how tax reform is implemented. Opinion is divided over whether the potential savings are significant. Third, there is no individual earnings uncertainty in our model. Consequently, there is no "precautionary" savings motive. Simulations undertaken by Engen and Gale (1997) and Engen, Gravelle, and Smetters (1997) suggest that this omission is more important *quan*titatively than it is *qual*itatively.[4] Finally, our analysis holds the supply of labor fixed: we defer discussion of the variable-work-effort case to a follow-up article, which will appear in a subsequent issue of *Economic Review*. As it turns out, the qualitative effects of tax reform in an economy with variable work effort differ little from those derived here—*provided that tax reform leaves the tax rate on labor income unchanged*.

We begin with a review of the basic features of the current tax system and three seemingly distinct, but actually equivalent, alternative types of flat-rate consumption tax. We then derive equations that characterize the savings and investment decisions of households and firms, conditional on the tax system. Each of these equations has a straightforward graphical interpretation. Using a set of diagrams, we analyze first the long-run, then the short-run impact of tax reform on output, consumption, investment, and interest rates. As already noted, the second article in this series achieves additional realism by extending the basic model to include variable work effort. Moreover, our second article uses simulations to explore the dynamic effects of tax reform in economies with capital adjustment costs and long-run growth effects.

ALTERNATIVE TAX SYSTEMS

Overview

The U.S. system of individual income taxes, payroll taxes, and corporate income taxes is exceedingly complex, involving numerous exemptions, deductions, credits, and carry-over provisions. It taxes different types of income at different rates, and the marginal tax rate applied to any given type of income may vary with income level. It makes no distinction between real capital gains and capital gains that simply reflect inflation. Similarly, depreciation allowances are based on nominal book values rather than replacement values.

We abstract from much of the complexity of the actual tax code. Thus, we assume that inflation is low enough that we can ignore its impact and conduct our analysis entirely in real terms. We allow wage income, corporate earnings, and interest and dividend income to be taxed at different rates but assume that the different marginal tax rates applied to these types of income are independent of the level of income. Each household feels free to buy and sell stock and other assets, but because households are assumed to be identical to one another, they never have occasion to do so. Consequently, there are never any realized capital gains. Firms' investment in plant and equipment is financed entirely from retained corporate earnings.[5] All corporate earnings not used to finance investment or pay taxes are distributed to households as dividends. Although these assumptions may seem extreme, for our purposes they simply strip the current tax system down to its essential features.

The alternative tax system that we analyze differs from the present system in two respects. First, it would replace the current hodgepodge of tax rates—under which some types of income are taxed more than once—with an integrated, flat-rate system of taxation. Second, it would base taxation on consumption rather than income.

There are three different versions of the flat-rate consumption tax: the national retail sales tax, the value-added tax (VAT), and the Hall–Rabushka tax (after which the Armey–Shelby flat tax was modeled).[6] Under a retail sales tax, each consumer good is taxed on its entire value at the time of final sale. No tax is collected on goods at intermediate stages of production or distribution. In contrast, a VAT collects a little piece of tax revenue at each stop along the production and distribution chain, based on the amount of value added to the good at that stop: under a VAT, firms pay tax on their sales less the sum of their purchases from other businesses. The Hall–Rabushka tax works in exactly the same way as a VAT, except each firm's employees are paid with pretax dollars, and it is the employees who write checks to the government for the taxes due on the wage component of value added. (The nonwage component of value added—corporate cash flow or sales less purchases from other businesses less wages—continues to be taxable to the firm. See Gentry and Hubbard 1997 for a nice discussion of what is included in nonwage value added.) Effectively, the Hall–Rabushka tax is a value-added tax where each worker is treated as an independent contractor. There is substantial controversy over which of these taxes would be easiest to implement, in practice (Slemrod 1996). However, for our purposes, all three are equivalent. For no better reason than that it is closest in *appearance* to the current tax system, we have chosen to model the Hall–Rabushka flat tax.

Details

We assume that output is produced from capital (plant and equipment) and labor, subject to a constant-returns-to-scale production technology (so that a doubling of all inputs into the production process doubles output).[7] The constant-returns-to-scale assumption allows us to measure all quantities on a per-worker basis. For example, we will use y to denote output per worker produced by the representative firm, n to denote hours of employment per worker, and k to denote capital per worker. Each period, a certain fraction, δ, of existing capital wears out and must be replaced if the capital stock is not to shrink. Net investment (the net change in the capital stock from one period to the next) will be denoted by Δk. We use w, R, and r to denote the real before-tax wage, the real before-tax interest rate, and the real after-tax interest rate, respectively; while g, b, and Δb denote real government purchases, the real stock of government bonds outstanding, and net new government indebtedness, all measured on a per-worker basis.

We make several simplifying assumptions. As noted above, in our model economy all capital investment is financed from retained earnings, and all other earnings are paid out either as taxes or as dividends. There is no role for government transfer payments in a world where all households are identical, so we will ignore them. Within each tax regime, tax rates are assumed constant through time. There is no uncertainty. Finally, tax reform is not announced in advance.

In our model of the current tax system, the government applies three different tax rates to three different types of income. Wage income, wn, is taxed at rate τ_w. Corporate profits, $y - wn - \delta k$, are taxed at rate τ_p. Any after-tax profits that are not used to finance net new investment are distributed as dividends and are taxed at the same rate, τ_d, as is interest income, Rb. Hence, after-tax dividends are $(1 - \tau_d)[(1 - \tau_p)(y - wn - \delta k) - \Delta k]$. In the United States, the average marginal federal tax rates on wages and corporate profits are each about 35 percent, while the tax rate on interest (and dividend) income is roughly 25 percent.[8]

Under our alternative tax plan, a single tax rate, τ, is applied to both wage income, wn, and corporate cash flow, $y - wn - \delta k - \Delta k$. Interest on newly-issued government debt is tax free. (To prevent a windfall gain to "coupon clippers," the interest on bonds that were issued *prior* to tax reform would have to remain taxable to recipients.) The U.S. Treasury estimates that implementing the Armey–Shelby version of the Hall–Rabushka tax system would require a 22.4 percent average marginal tax rate on labor income (Auerbach 1996). Replacing the revenue from the current federal payroll tax would bring this tax rate up to a level roughly comparable to the rates of wage and profit taxation under the current system—that is, approximately 35 percent.[9]

We can use the government budget constraint to see the connection between the current tax system and our flat-rate alternative. In our stylized model of the current system, the government budget constraint is given by

$$(1) \quad \Delta b = g + rb - \{\underbrace{\tau_w wn}_{\text{Wage taxes}} + \underbrace{\tau_p(y - wn - \delta k)}_{\text{Profits taxes}} + \underbrace{\tau_d[(1 - \tau_p)(y - wn - \delta k) - \Delta k]}_{\text{Dividend taxes}}\}.$$

This equation simply says that the government must issue more debt whenever its expenditures (on goods and services, and net interest) exceed the revenue it receives from taxing wage, profit, and dividend income. In an economy with a flat-rate consumption tax, in contrast, the government budget constraint takes the form

$$\begin{aligned}(2) \quad \Delta b &= g + rb - [\underbrace{\tau wn}_{\text{Wage taxes}} + \underbrace{\tau(y - wn - \delta k - \Delta k)}_{\text{Cash flow taxes}}] \\ &= g + rb - \tau(y - \delta k - \Delta k) \\ &= g + rb - \tau(c + g),\end{aligned}$$

where $c = y - \delta k - \Delta k - g$ and denotes real consumption expenditures. Note that imposing a uniform tax on wage income and corporate cash flow is equivalent to taxing the sum of household and government spending on goods and services.

We stated above that one can think of our alternative tax plan as being two steps removed from the current tax system. The first step takes us from the current system to an integrated, flat-rate income tax. In our model, this step is accomplished by setting $\tau_d = 0$ (eliminating the double taxation of corporate earnings) and $\tau_w = \tau_p \equiv \tau'$. Equation 1 reduces to

$$(1') \qquad \Delta b = g + rb - \tau'(y - \delta k).$$

Note the similarity between Equation 1′ and the second line of Equation 2. To complete the move from our current income tax system to a flat-rate consumption tax requires only the additional step of allowing firms to deduct ***all*** purchases of plant and equipment, not just depreciation on ***existing*** plant and equipment, before calculating their tax liability.[10]

UTILITY AND PROFIT MAXIMIZATION

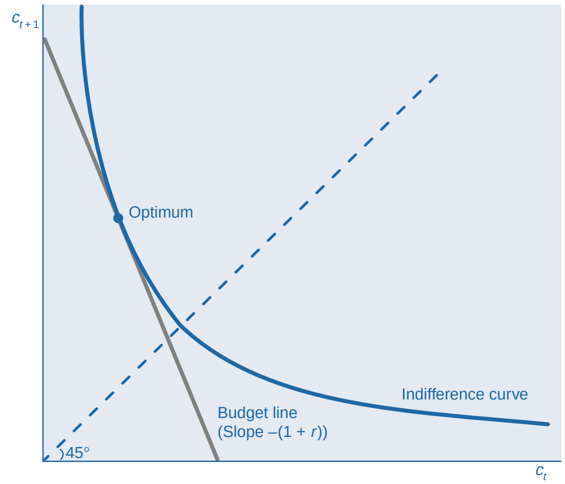
In this section, we discuss the implications that utility and profit maximization have for the relationship between the variables in our model economy. This discussion lays the necessary groundwork for all of our subsequent analysis. To keep the model as simple as possible, we assume that there are no capital adjustment costs and that the supply of labor is exogenously fixed. These assumptions are relaxed in the sequel to this article.

The Household Savings Decision

The optimality conditions of the representative household equate the rate at which the household is willing to trade one good for another (the ***marginal rate of substitution*** between the goods) to the relative market prices of the goods. For example, suppose that we denote by $MRS(c_t, c_{t+1})$ the number of units of consumption at time $t + 1$ required to compensate the representative household for the loss of one unit of consumption at time t. Moreover, suppose that r_t denotes the after-tax interest rate at time t (so that $r_t = (1 - \tau_d)R_t$ under the current tax system and $r_t = R_t$ after tax reform). Then only when $1 + r_t = MRS(c_t, c_{t+1})$ will the household's allocation of consumption across time be optimal.[11] Graphically, the household will allocate consumption so as to be at a point of tangency between its intertemporal budget constraint and one of its indifference curves (*Figure 1*).

It is standard to assume that the marginal rate of substitution is decreasing in its first argument and increasing in its second argument. In the present case, this condition means that households tend to prefer smooth consumption paths to uneven ones. (The indifference curves in Figure 1 are convex to the origin.) Also, households respond to an increase in wealth by demanding more current and future

Figure 1
Choosing Between Current and Future Consumption



consumption. (A parallel outward shift in the household budget line shifts the point of tangency between the budget line and the household's indifference curves to the northeast.) It is also standard to assume that $MRS(c_t, c_{t+1}) = 1 + \rho$, for some fixed $\rho > 0$, whenever $c_t = c_{t+1}$. (The representative household's indifference curves have slope $-(1 + \rho)$ where they cross a 45° line extending out from the origin.) The parameter ρ is the household's ***pure rate of time preference.***

From our assumptions about household preferences, it follows immediately that $MRS(c_t, c_{t+1}) > 1 + \rho$ if, and only if, $c_{t+1} > c_t$. However, we have already seen that an optimizing household will equate the marginal rate of substitution between current and future consumption to one plus the after-tax interest rate. Hence,

$$(3) \qquad \begin{aligned} & c_{t+1} > c_t \\ & \Leftrightarrow MRS(c_t, c_{t+1}) > 1 + \rho \\ & \Leftrightarrow r_t > \rho. \end{aligned}$$

In words, consumption will be rising through time if, and only if, the real after-tax interest rate exceeds the pure rate of time preference. Intuitively, a high after-tax rate of return on saving is needed to induce households to defer consumption.

The Business Investment Decision

The optimality conditions that characterize the representative firm's investment decision are different under a consumption tax than they are under the current income tax system. We look first at the income tax case, then turn our attention to the consumption tax case.

Investment Under an Income Tax. Under an income tax system, after-tax dividends are

$$(4)\qquad (1 - \tau_d)[(1 - \tau_p)(y - wn - \delta k) - \Delta k].$$

Hence, increasing period-t investment by one unit requires that period-t dividends be cut by $(1 - \tau_d)$ units. If used to purchase government bonds, these $(1 - \tau_d)$ units of period-t output would yield $(1 + r_t)(1 - \tau_d)$ units of output in period $t + 1$. The firm should continue to increase its capital investment as long as it can give shareholders a better marginal return than they would receive under this fallback strategy.

With one additional unit of capital available in period $t + 1$, the firm's production will be higher than would otherwise have been the case, but so will its depreciation costs. On net, taxable profits rise by $MP_k - \delta$, where MP_k denotes the marginal product of capital—the increment to production from an additional unit of capital. After-tax profits rise by $(1 - \tau_p)(MP_k - \delta)$. Moreover, because it increased capital investment in period t, the firm will be able to avoid one unit of capital investment in period $t + 1$. Thus, altogether, the firm will be able to increase period $t + 1$ dividends by $1 + (1 - \tau_p)(MP_k - \delta)$ units of output if it increases current investment by one unit. Of course, only the fraction $(1 - \tau_d)$ of these dividends will be available to shareholders after taxes.

Summarizing, when it increases period-t investment by one unit, the firm deprives its shareholders of $(1 + r_t)(1 - \tau_d)$ units of output in period $t + 1$ and, in exchange, gives them $(1 - \tau_d)[1 + (1 - \tau_p)(MP_k - \delta)]$ units of output. A profit-maximizing firm will expand investment until the marginal return to investment just equals the marginal cost of investment:

$$(1 - \tau_d)[1 + (1 - \tau_p)(MP_k - \delta)] = (1 + r_t)(1 - \tau_d),$$

or, equivalently,

$$(5)\qquad (1 - \tau_p)(MP_k - \delta) = r_t.$$

Thus, under the current income tax system, firms invest up to the point where the marginal product of capital net of depreciation and incremental profits taxes equals the after-tax interest rate.

Investment Under a Consumption Tax. Under the Hall–Rabushka version of the flat-rate consumption tax, after-tax dividends are

$$(4')\qquad (1 - \tau)(y - wn - \delta k - \Delta k).$$

Hence, the opportunity cost of capital investment is $(1 - \tau)$ units of period-t output or, equivalently, $(1 - \tau)(1 + r_t)$ units of output in period $t + 1$. The marginal, after-tax return to capital investment is $(1 - \tau)(MP_k + 1 - \delta)$ units of output in period $t + 1$. (The firm has MP_k additional units of newly produced output to sell in period $t + 1$, plus used equipment worth $1 - \delta$ units of output.) Therefore, the marginal return to investment just equals the marginal cost of investment when

$$(1 - \tau)(MP_k + 1 - \delta) = (1 - \tau)(1 + r_t)$$

or, equivalently, when

$$(5')\qquad MP_k - \delta = r_t.$$

Under the consumption tax, firms invest up to the point where the net-of-depreciation marginal product of capital equals the real interest rate.

The Output Market

Of course, business investment and household savings decisions are not independent of one another. They are linked by the requirement that the sum of consumption, investment, and government purchases equals the total amount of output produced. Formally, we must have $f(k, n_0) = c + \delta k + \Delta k + g$, where $f(\bullet, \bullet)$ gives the amount of output produced per worker as a function of the amount of capital per worker and the number of hours of employment per worker (held fixed at n_0). Turning this equation around,

$$(6)\qquad \Delta k = f(k, n_0) - c - g - \delta k.$$

Hence,

$$(7)\qquad \Delta k > 0 \Leftrightarrow c < f(k, n_0) - g - \delta k.$$

Equation 7 simply states that the capital stock will increase when consumption is low relative to production (net of government purchases and depreciation) and will decrease when consumption is high relative to production. In the former case, there is more than enough output available, after deducting household and government consumption, to replace plant and equipment as it wears out. In the latter case, so much output is being consumed that firms are unable to replace worn-out plant and equipment. One can think of $f(k, n_0) - g - \delta k$ as being the level of consumption that is sustainable, given the capital stock and the level of government purchases.

The Labor Market

Finally, consider the labor market. As noted above, we assume that the supply of labor is fixed at $n = n_0$. The demand for labor is determined by profit maximization. The representative firm will demand additional labor as long as the incremental labor adds more to revenues (through increased production) than it adds to costs (through increased wages). Under either tax system, the increment to revenues is simply the marginal product of labor—denoted by MP_n—and the increment to costs is simply the real wage, w. Hence, profit maximization implies that

$$MP_n = w. \tag{8}$$

The marginal product of labor must equal the real wage.

THE EFFECTS OF TAX REFORM: THE LONG RUN

In this section, we develop a set of diagrams that summarizes the optimality and market-clearing conditions we derived above. We use this set of diagrams to analyze the long-run impact of a flat-rate consumption tax on consumption, the capital stock, and interest rates. We also consider the long-run impact of tax reform on wages and the stock market. All discussion of the *transition* from one long-run equilibrium to another is deferred until later in the article.

The Demand for Capital and the Long-Run Supply of Capital

Our model economy abstracts from any source of sustained growth, such as technological change. Consequently, the long-run equilibrium in our model will be characterized by a constant level of consumption and a constant capital stock. We already know (from Equation 3) that households will be content with a constant level of consumption if, and only if, the after-tax interest rate equals the pure rate of time preference: $r_t = \rho$. Thus, there is only one after-tax interest rate consistent, in the long run, with the optimality condition that governs household saving decisions: as shown in the bottom panel of Figure 2, the long-run capital-supply curve is horizontal at the pure rate of time preference.

The capital-demand curve differs depending on the tax regime. According to the investment-optimality condition for a firm subject to a corporate income tax (Equation 5), the real after-tax interest rate that will just induce the representative firm to hold a given quantity of capital is $(1 - \tau_p)(MP_k - \delta)$. Assuming that the marginal product of capital is decreasing in the capital stock, this optimality condition defines a downward-sloping relationship between r and k. See the gray line plotted in the bottom panel of Figure 2.

Under a flat-rate consumption tax, the investment-optimality condition is Equation 5′. It says that the representative firm will add to its capital stock up to the point where net-of-depreciation marginal product of capital just equals the after-tax interest rate. So now the critical interest rate is $MP_k - \delta$ rather than $(1 - \tau_p)(MP_k - \delta)$. The capital demand curve is still downward sloping, but it is proportionately higher than the capital demand curve under the income tax. See the blue line plotted in the bottom panel of Figure 2.

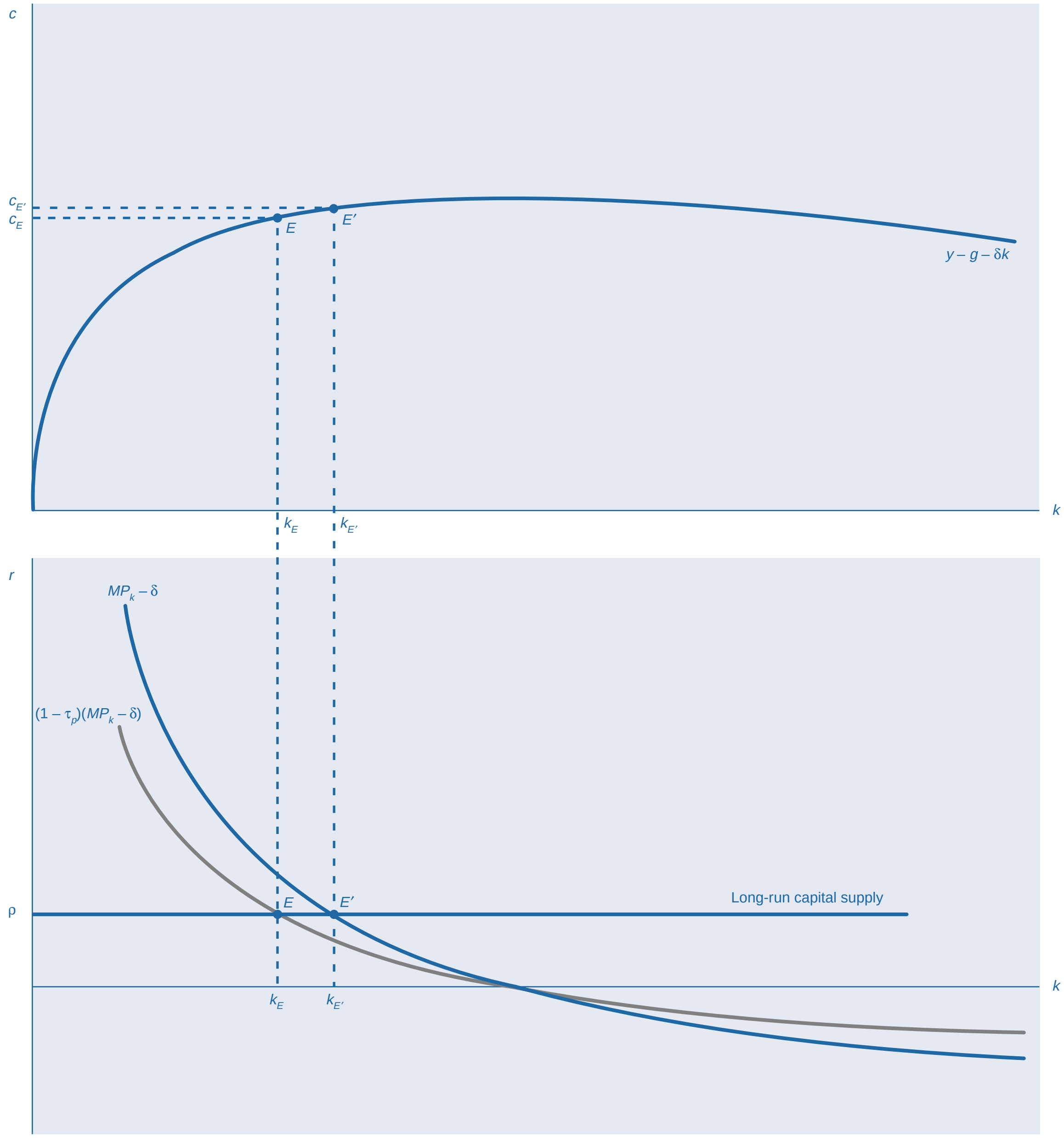
The Long-Run Impact of Tax Reform

Long-run equilibrium in the capital market occurs where the capital-demand and capital-supply schedules intersect. In the bottom panel of Figure 2, this intersection occurs at point E in the economy with a corporate income tax, and at point E' in the economy with a flat-rate consumption tax. The after-tax interest rate is the same in the two economies, but the economy with the consumption tax has a larger steady-state capital stock—for realistic parameter values, 29 percent higher. (See the box titled "A Numerical Example" on page 37.) A higher capital stock means more output—roughly 9 percent more than under an income tax.[12]

What of the *pre*tax interest rate? In steady state under either tax system, we know that $r = \rho$. Under a consumption tax, households do not pay tax on their interest income. So $R = r = \rho$ under a consumption tax. Under an income tax, in contrast, interest is taxed at rate τ_d, so $R = r/(1 - \tau_d) = \rho/(1 - \tau_d)$. Thus, the steady-state pretax interest rate is lower (by about 25 percent) under a consumption tax than under an income tax.

What of the stock market value of the typical firm? Under a consumption tax, each additional unit of capital investment costs shareholders $1 - \tau$ units of current after-tax dividends (Equation 4′). Hence, each unit of capital is worth $1 - \tau$ units of consumption at the margin, and the real value of the firm is $(1 - \tau)k$. Over time, this value approaches $(1 - \tau)k_{E'}$, where $k_{E'}$ is the steady-state capital stock. Under an income tax, each unit of capital investment costs shareholders $1 - \tau_d$ units of current after-tax dividends (Equation 4). So the real value of the firm is $(1 - \tau_d)k_E$ in steady state. Whether tax reform ultimately raises or ultimately lowers stock prices is, in general, ambiguous. The capital stock is clearly higher after reform, but the tax rate applied to corporate cash flow after reform might also very well be higher than the current tax rate on dividends. (Recall that $\tau_d \approx 0.25$,

Figure 2
Pre-Reform and Post-Reform Long-Run Equilibria



whereas replacing the revenues from the current income and payroll tax systems, while allowing for some initial amount of wage income to be tax exempt, requires $\tau \approx 0.35$.) As noted above, realistic parameter values suggest that the steady-state capital stock is almost 29 percent higher under a consumption tax than under an income tax. Consequently, it is reasonable to expect that real stock prices would ultimately increase by a little less than 12 percent as a result of tax reform.

What of the pretax wage? From profit maximization, we know that the wage rate equals the marginal product of labor (Equation 8). With a constant-returns-to-scale production technology, the marginal product of labor depends only on the capital/labor ratio. Since the steady-state capital stock is higher under a consumption tax than under an income tax, the same must be true of steady-state labor productivity and the steady-state real wage. For realistic parameter values, the real wage rises by about 9 percent in the long run.

We know that consumption is constant in the long-run equilibrium of our model economy —but constant at what level? We can use the top panel of Figure 2 to find out. This panel shows a plot of the function $f(k, n_0) - g - \delta k$, which we know from our discussion of the output-market clearing condition (Equation 7) is the formula for the maximum sustainable level of consumption. In plotting sustainable consumption, we have assumed that capital is necessary for producing output ($f(0, n_0) = 0$) and that the net marginal product of capital ($MP_k - \delta = f_1(k, n_0) - \delta$) is positive at low levels of capital and decreasing in the quantity of capital. Consequently, the $y - g - \delta k$ curve has vertical intercept $-g$ and an inverted-U shape. It attains its maximum when the marginal product of capital equals the depreciation rate ($f_1(k, n_0) - \delta = 0$).

To find the steady-state level of consumption graphically, we need only move upward from points E and E' in the lower panel of Figure 2 to the corresponding points along the curve plotted in the upper panel. Since the $y - g - \delta k$ curve is necessarily upward sloping over the relevant range, a higher steady-state capital stock implies a higher steady-state level of consumption. In the diagram, the steady-state level of consumption under the consumption tax ($c_{E'}$) is greater than the steady-state level of consumption under the income tax (c_E).

To review, the key difference between an income tax and a consumption tax is that the former does not allow firms to expense their capital investment. (Compare Equations 4 and 4′, or Equations 1′ and 2.) Consequently, the trade-off between current dividends and future dividends is distorted under an income tax: shareholders have a bias in favor of current dividends that is lacking under a consumption tax. (Compare Equations 5 and 5′.) This bias drives firms to demand less capital, at any given after-tax interest rate, than they would under a consumption tax. Because the after-tax interest rate must, in steady state, equal the pure rate of time preference, the capital stock ends up at a lower level under an income tax. Since the steady-state capital stock is lower under an income tax, so are steady-state output and steady-state consumption.

THE EFFECTS OF TAX REFORM: DYNAMICS

So far we have considered only the long-run effects of tax reform. If we are interested in the path of the economy *between* steady states, we must modify our graphical apparatus. Fortunately, the required changes are not large—our new diagram is quite similar to Figure 2. However, because interactions between consumption and capital are important in the short run, the focus of our analysis shifts away from the bottom panel of the diagram and to the top panel.[13] We demonstrate that consumption at first must decline following tax reform to make room for increased investment. The stock market is also likely to decline, while interest rates are likely to rise.

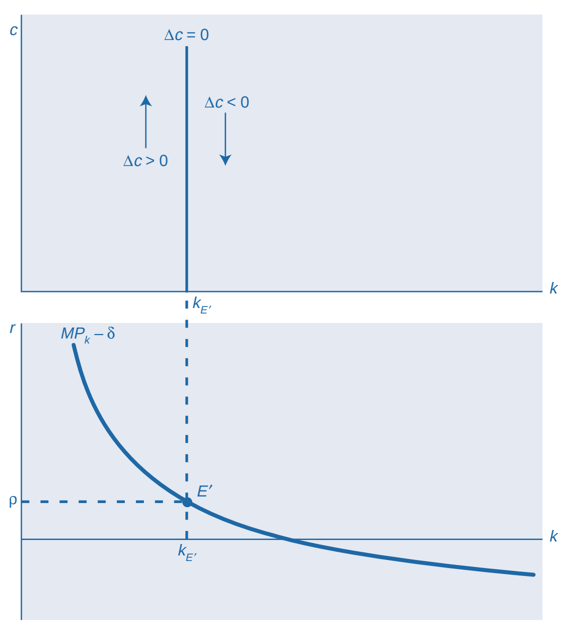
The Phase Diagram

Since our interest is in how the U.S. economy would evolve following tax reform, we analyze the short-run dynamics of an economy in which the government relies on a flat-rate consumption tax to meet its revenue needs. Consider first the dynamics of consumption. From Equation 3 (the optimality condition for household saving) we know that households are willing to defer consumption if, and only if, the after-tax interest rate exceeds the pure rate of time preference. However, from Equation 5′ (the optimality condition for investment) we know that for the representative firm to be willing to hold its capital stock, the after-tax interest rate must equal the net marginal product of capital. Putting these two conditions together, we find that consumption will be rising over time if, and only if, the net marginal product of capital exceeds the pure rate of time preference:

$$(9) \qquad \Delta c > 0 \Leftrightarrow MP_k - \delta > \rho.$$

The bottom panel of Figure 3, much like the bottom panel of Figure 2, shows a plot of the net marginal product of capital. Clearly, the net marginal product of capital exceeds the time-

Figure 3
Short-Run Dynamics of Consumption



preference rate ρ if, and only if, the capital stock is less than the steady-state capital stock, $k_{E'}$. Hence,

$$\Delta c > 0 \Leftrightarrow k < k_{E'}. \quad (10)$$

In words, consumption will be increasing over time if, and only if, the capital stock falls short of its steady-state level. Intuitively, if capital is scarce relative to labor, then the return on new capital investment will be high, inducing households to sacrifice some current consumption in exchange for higher future consumption.

In the top panel of Figure 3 we put an upward-pointing arrow to the left of $k_{E'}$, reflecting the fact that consumption will be increasing over time whenever $k < k_{E'}$. Similarly, we place a downward-pointing arrow to the right of $k_{E'}$. At $k_{E'}$ itself, we put a vertical line, labeled $\Delta c = 0$, to indicate that here consumption tends to neither rise nor fall.

We turn now to the dynamics of the capital stock. Equation 7 says that the capital stock will tend to fall, over time, at points above the curve $y - g - \delta k$ in the top panel of Figure 2. At points below the curve, the capital stock will tend to increase. The intuition is that a level of consumption that is high relative to the capital stock can be achieved only by not replacing capital equipment as it wears out. If, on the other hand, consumption is *low* relative to the capital stock, then there is more than enough output left over (after meeting the demands of households and the government) to replace worn-out capital, and the capital stock rises over time. In Figure 4, the $y - g - \delta k$ curve is relabeled as $\Delta k = 0$, and arrows are placed above and below it, pointing to the left and right, respectively.

Figure 4
Short-Run Dynamics of Capital

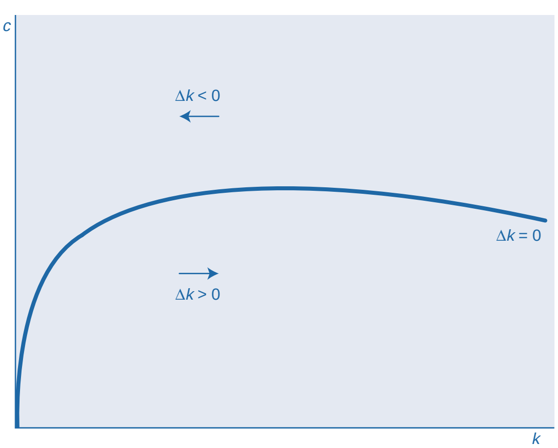


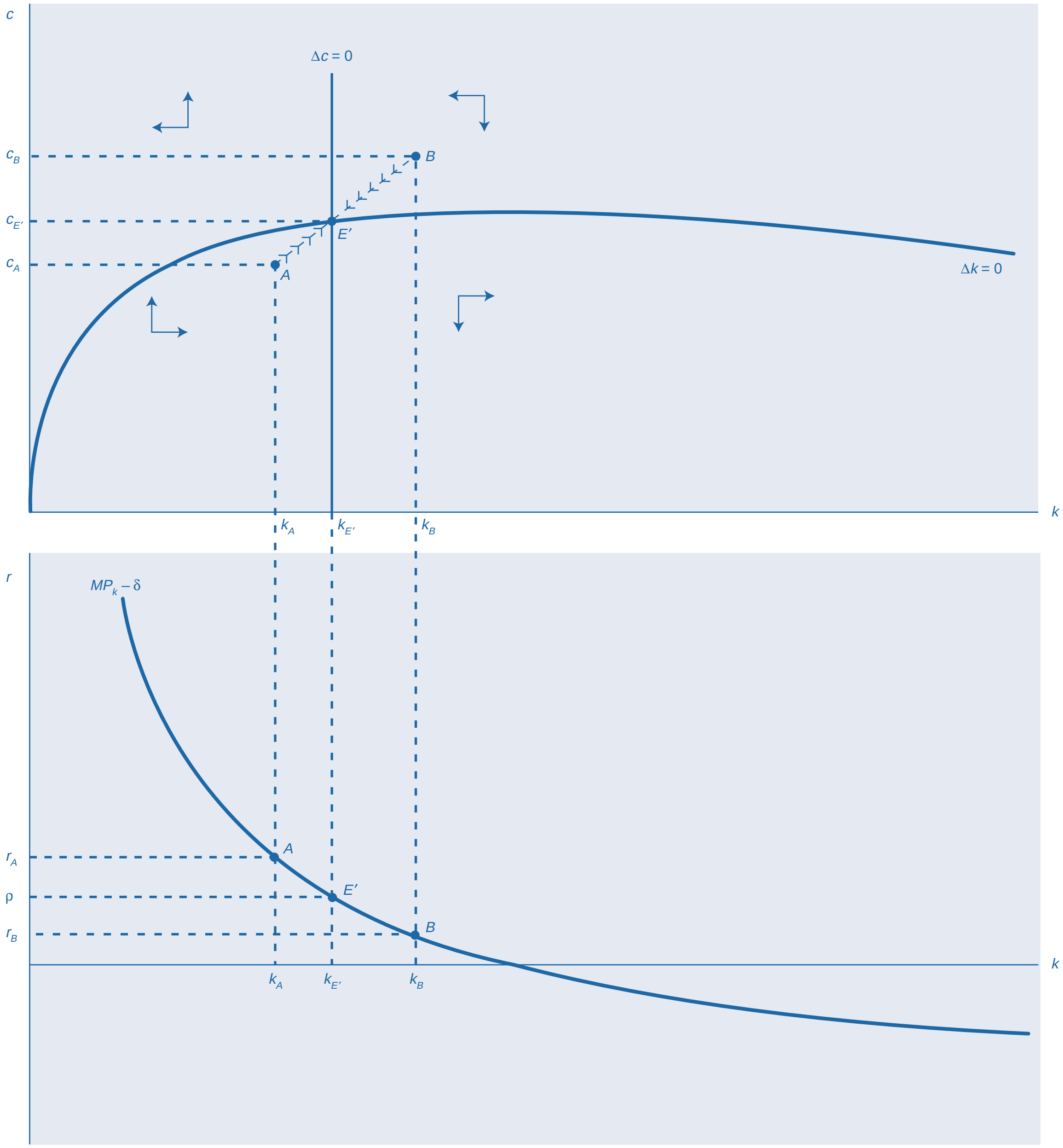
Figure 5 combines Figures 3 and 4. In the figure's top panel, arrows show the directions of consumption and capital movement for different levels of consumption and capital. Its bottom panel shows how the after-tax interest rate varies with the capital stock. As in Figure 2, the economy's unique steady state is point E'. In this steady state, $r = \rho$, $k = k_{E'}$, and $c = c_{E'} = f(k_{E'}, n_0) - g - \delta k_{E'}$. Point E' is called a ***saddle-path equilibrium.*** For each initial capital stock, there is a unique level of consumption such that the economy will approach E'. Any other initial consumption level would put the economy on either an explosive or an implosive path—a path that cannot be optimal.[14] In the diagram, if the economy starts at some arbitrary capital stock $k_A < k_{E'}$, then households will choose consumption level $c_A < c_{E'}$ and the economy will follow the dashed path from point A toward point E'. As the capital stock increases, the after-tax interest rate falls from r_A toward ρ. Similarly, if the economy starts at some capital stock $k_B < k_{E'}$, then households will choose consumption level $c_B > c_{E'}$ and the economy will follow the dashed path from point B toward point E'. The interest rate rises from r_B toward ρ.

The Effects of Tax Reform: The Short Run

Consider an economy in steady state under an income tax. Suddenly, the income tax is replaced with a consumption tax.[15] We know that the economy starts at point E in the upper panel of Figure 6, and eventually ends up at point E'. What happens along the way? Our phase diagram gives us the answer. We know that point E' is a saddle-point equilibrium under the new tax system: there is a unique path for consumption and the capital stock that simultaneously satisfies all of the utility and profit maximization conditions and that is neither explosive nor implosive. This path runs through E', and movements along the path are governed by the set of directional arrows depicted in the figure. For any given initial capital stock, households will choose the level of consumption that puts the economy on this convergent path.

In the upper panel of Figure 6, the economy jumps downward from point E to point A the instant that tax reform is put into effect. Intuitively, the after-tax return on capital jumps upward from ρ to $r_A = \rho/(1 - \tau_p)$ with the switch to a consumption tax. (In the lower panel of Figure 6, the economy jumps upward from point E to point A.) The higher marginal return to capital means that households are willing to

Figure 5
Combined Short-Run Dynamics

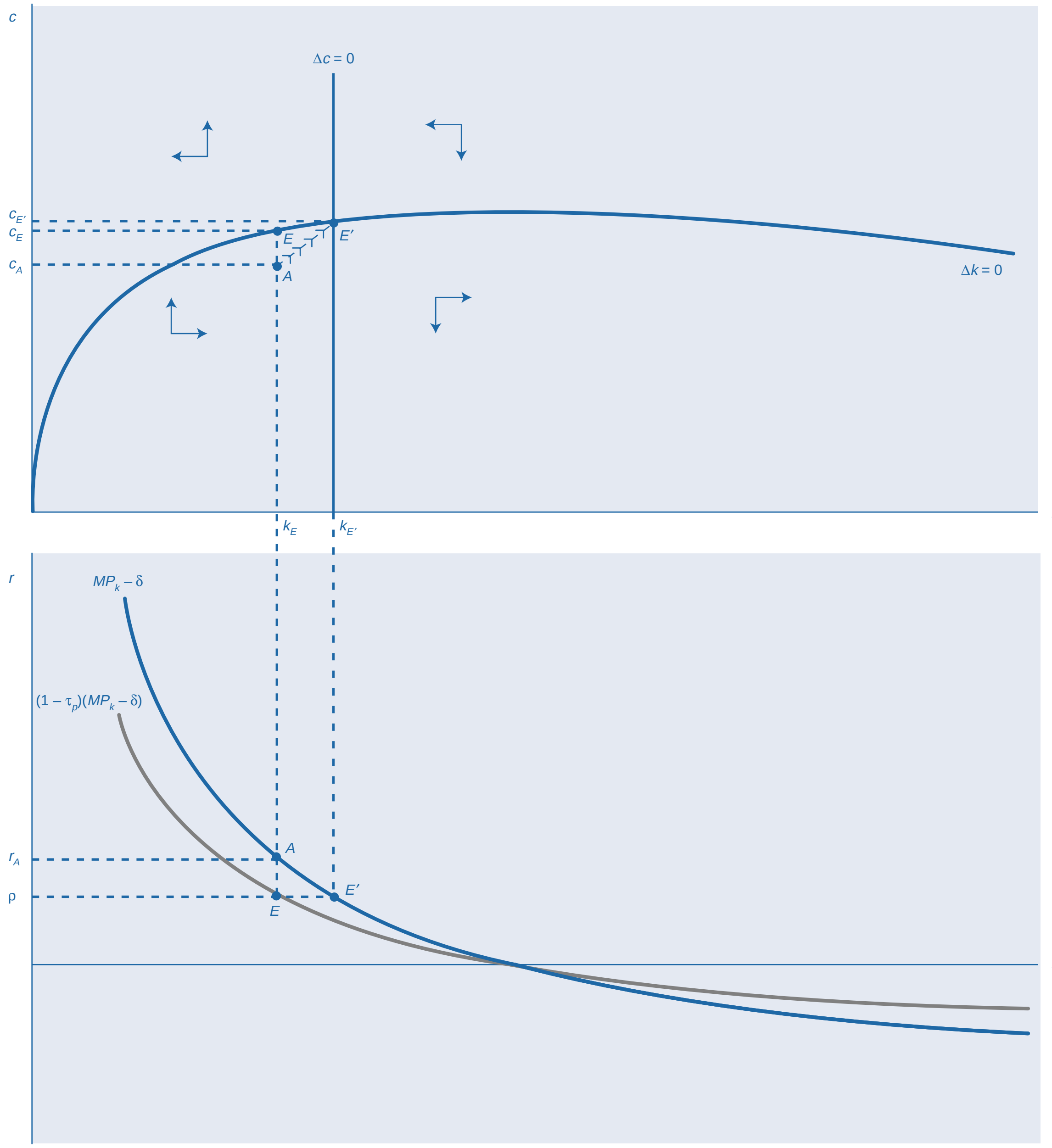


take a cut in dividends (reduce their consumption) in order that firms may finance the acquisition of additional capital through higher retained earnings. As the capital stock gradually expands, the economy moves up the dashed saddle path from point A toward point E' in the upper panel of Figure 6, and down the capital demand curve from point A to point E' in the lower panel.

The *pre*tax interest rate, R, is subject to conflicting influences in the short run. In the initial steady state, $R = \rho/(1 - \tau_d)$. Immediately after reform is implemented, $R = r_A = \rho/(1 - \tau_p)$. Thus, the before-tax interest rate will fall if, and only if, $\tau_p < \tau_d$. In fact, though, $\tau_p \approx 0.35 > 0.25 \approx \tau_d$. Hence, the pretax interest will likely rise by about 15 percent with the imposition of a consumption tax before gradually declining to its new steady-state level (ρ).

The immediate impact of tax reform on stock prices also depends upon relative tax rates. The steady-state level of stock prices under an income tax is $(1 - \tau_d)k_E$. Immediately following the move to a consumption tax, the level of stock prices is $(1 - \tau)k_E$. If, as argued above, $\tau_d \approx 0.25$ and $\tau \approx 0.35$, then stock prices

Figure 6
Dynamic Response to Tax Reform



will fall by about 13 percent upon the implementation of tax reform. However, this result is sensitive to changes in our assumptions about the features of tax reform. For example, if existing payroll taxes are kept in place (so that the consumption tax need only replace the revenue from the current income tax), then the U.S. Treasury estimates that $\tau \approx 0.224$. In this case, stock prices would actually jump upward slightly following tax reform. In any event, following their initial jump, stock prices vary with the capital stock, gradually rising toward $(1 - \tau)k_E$.

Finally, the pretax wage rate is linked to the capital stock via the marginal product of labor (Equation 8). Since the capital stock doesn't jump, neither does the wage rate. As the capital stock gradually increases, so does the wage rate. Whether the *after*-tax wage rate jumps upward or downward depends entirely on the size of τ_w relative to τ. Under our base-case scenario, these tax rates are equal. Consequently, the after-tax wage, like the pretax wage, does not initially move.

An illustrative simulation of the effects of fundamental tax reform is presented in the box.

SUMMARY AND CONCLUDING REMARKS

Table 1 summarizes our principal findings. The first column of the table shows the immediate impact that the adoption of a consumption tax can be expected to have on each of several variables. The second column shows the long-run impact of tax reform. Ordinarily, when we tax a good or activity, we expect to see *less* of it in the marketplace. However, a tax on consumption causes the economy to achieve a *higher* level of consumption, in the long run, than would be observed under an income tax. The resolution of this paradox is that society accumulates greater real wealth under a consumption tax than it does under an income tax—an accumulation that is made possible because the *initial* effect of the consumption tax is to reduce consumption.

Is the eventual increase in consumption worth the initial decline? In the simple model economy examined here, the answer is unambiguously yes. Because the supply of labor is fixed in our model economy, the only component of the income tax that is distortionary is the profits tax: it reduces the demand for capital at any given after-tax interest rate. Because the consumption tax eliminates this distortion, it necessarily raises social welfare.

The real world is obviously more complicated than our model economy. Most pertinently, household labor supply is not exogenously fixed: any tax on labor income distorts households' labor–leisure choices. If moving from an income tax to a consumption tax significantly worsens this labor-market distortion, it may reduce social welfare despite the fact that, at the same time, it eliminates an investment-saving distortion.

Fortunately, it is relatively easy to explain how the results obtained here would change, qualitatively, if the supply of labor was endogenous. We address this issue in Part 2 of our article. Part 2 also considers the sensitivity of our results to capital adjustment costs and to a specification of the production function that has each firm's output depend on the *aggregate* capital stock as well as its own.

Table 1
Impact of Fundamental Tax Reform in the Basic Model

Variable	Immediate impact	Long-run impact
Output, capital	0	+
Consumption	–	+
After-tax interest rate	+	0
Pretax interest rate	+[a]	–
Stock market	–[b]	+[c]
Real wage	0	+

NOTES: [a] Assumes that profits are currently taxed more heavily than is interest.
[b] Assumes that the new tax rate on corporate cash flow will exceed the current tax rate on dividends.
[c] Assumes that $(1 - \tau)k_{E'}$ exceeds $(1 - \tau_d)k_E$.

NOTES

1 Howitt and Sinn (1989) undertake a more sophisticated analysis within a similar framework.

2 For simulation exercises that examine the impact of tax reform on different age groups, see Auerbach and Kotlikoff (1987) and Auerbach (1996). Feenberg, Mitrusi, and Poterba (1997) present a careful analysis of the impact of tax reform on the distribution of consumption. Sarkar and Zodrow (1993) discuss windfall gains and losses resulting from tax reform and how they might be mitigated. Monetary policy affects the distribution of wealth primarily through unanticipated inflations, which benefit debtors at the expense of lenders, and unanticipated deflations, which have the opposite effect.

3 Hines (1996) contains a general discussion of complications that arise in an open-economy setting. Mendoza and Tesar (1995) construct a formal model.

4 In otherwise identical models, the inclusion of a precautionary savings motive cuts the impact of tax reform on output, consumption, and the capital stock roughly in half.

5 According to the Federal Reserve Board's flow of funds accounts for the nonfarm, nonfinancial corporate business sector, internal funds averaged 94.8 percent of capital expenditures over the ten-year period from 1985 through 1994. In contrast, credit market borrowing averaged only 32.5 percent of capital expenditures.

6 For more detailed descriptions of these alternative approaches, see Koenig and Taylor (1996), Hall and Rabushka (1996), Metcalf (1996), and Moore (1996).

7 We ignore human capital and the associated investment in education and training.

8 The statutory tax rate on corporate profits is 35 percent. The tax rate on wage income can be obtained by adding 14 percent (representing federal payroll taxes) to the 22 percent marginal income tax rate reported in Auerbach (1996). The actual tax rates on household interest and dividend income are 22 percent and 27 percent, respectively (Auerbach 1996). For comparison, Mendoza, Razin, and Tesar (1994) obtain an estimate of 32 percent for the U.S. corporate capital income tax rate and an estimate of 34 percent for the tax rate on U.S. labor income. Triest (1996) estimates a 38 percent average marginal tax rate on labor income. Personal income tax deductions and exemptions that are excluded from our model (and that would be eliminated under most tax-reform proposals) account for the relatively low revenue yield of the current system, despite high marginal tax rates.

9 Triest's (1996) estimate of the required flat-tax rate (39 percent) is somewhat higher than ours, but not much different from his own estimate of the current average marginal tax rate on labor income (38 percent).

10 The alert reader will have noted that the transition can

be completed in a single step by setting $\tau_p = 0$ and $\tau_w = \tau_d \equiv \tau'$. However, this approach breaks down in the real world, where not every firm is able to finance its capital investment out of its own retained earnings.

[11] If $1 + r_t > MRS(c_t, c_{t+1})$, then it will be advantageous to the household to reduce its period-*t* consumption by one unit and buy a bond. Principal and interest on the bond (received in period $t + 1$) will be more than enough to compensate the household for the reduction in c_t. Similarly, if $1 + r_t < MRS(c_t, c_{t+1})$, then it will be advantageous to the household to *sell* a bond from its portfolio and use the proceeds to *increase* its period-*t* consumption by one unit.

[12] These estimates are meant to convey no more than the likely order of magnitude of the economy's response to tax reform. They are sensitive to the assumed profits tax rate and to the assumed capital-elasticity of output. Other studies have generally reported a slightly weaker baseline output response.

[13] Moreover, we implicitly switch to a continuous-time version of the model described above.

[14] Think of a marble on a saddle. The surface of the saddle looks like a U when viewed from the side and like an inverted U when viewed from either end. The point in the middle of the saddle is the steady state. A marble placed precisely at this point will remain stationary. In principle, a marble placed at the exact middle of one of the saddle's ends will move toward the steady state rather than fall off the saddle on either side.

[15] For a gradual reform analysis, see Howitt and Sinn (1989).

REFERENCES

Auerbach, Alan J. (1996), "Tax Reform, Capital Allocation, Efficiency, and Growth," in *Economic Effects of Fundamental Tax Reform*, eds. Henry J. Aaron and William G. Gale (Washington, D.C.: Brookings Institution Press), 29–82.

Auerbach, Alan J., and Laurence J. Kotlikoff (1987), *Dynamic Fiscal Policy* (Cambridge, England: Cambridge University Press).

Engen, Eric M., and William G. Gale (1997), "Consumption Taxes and Saving: The Role of Uncertainty in Tax Reform," *American Economic Review Papers and Proceedings* 87 (May): 114–19.

Engen, Eric M., Jane Gravelle, and Kent Smetters (1997), "Dynamic Tax Models: Why They Do the Things They Do," *National Tax Journal* 50 (September): 657–82.

Feenberg, Daniel R., Andrew W. Mitrusi, and James M. Poterba (1997), "Distributional Effects of Adopting a National Retail Sales Tax," in *Tax Policy and the Economy*, ed. James M. Poterba (Cambridge, Mass.: MIT Press), 49–89.

Gentry, William M., and R. Glenn Hubbard (1997), "Distributional Implications of Introducing a Broad-Based Consumption Tax," in *Tax Policy and the Economy*, ed. James M. Poterba (Cambridge, Mass.: MIT Press), 1–47.

Hall, Robert E., and Alvin Rabushka (1996), "The Flat Tax: A Simple, Progressive Consumption Tax," in *Frontiers of Tax Reform*, ed. Michael J. Boskin (Stanford, CA: Hoover Institution Press), 27–53.

Hines, James R. (1996), "Fundamental Tax Reform in an International Setting," in *Economic Effects of Fundamental Tax Reform*, eds. Henry J. Aaron and William G. Gale (Washington, D.C.: Brookings Institution Press), 465–502.

Howitt, Peter, and Hans-Werner Sinn (1989), "Gradual Reforms of Capital Income Taxation," *American Economic Review* 79 (March): 106–24.

Koenig, Evan F., and Lori L. Taylor (1996), "Tax Reform: Is the Time Right for a New Approach?" Federal Reserve Bank of Dallas *Southwest Economy*, Issue 1, January/February, 5–8.

Mendoza, Enrique G., and Linda L. Tesar (1995),"Supply-Side Economics in a Global Economy," International Finance Discussion Paper no. 507 (Washington, D.C.: Board of Governors of the Federal Reserve System, March).

Mendoza, Enrique G., Assaf Razin, and Linda L. Tesar (1994), "Effective Tax Rates in Macroeconomics: Cross-Country Estimates of Tax Rates on Factor Incomes and Consumption," *Journal of Monetary Economics* 34 (December): 297–323.

Metcalf, Gilbert E. (1996),"The Role of a Value-Added Tax in Fundamental Tax Reform," in *Frontiers of Tax Reform*, ed. Michael J. Boskin (Stanford, CA: Hoover Institution Press), 91–109.

Moore, Stephen (1996),"The Economic and Civil Liberties Case for a National Sales Tax," in *Frontiers of Tax Reform*, ed. Michael J. Boskin (Stanford, CA: Hoover Institution Press), 110–20.

Sarkar, Shounak, and George R. Zodrow (1993), "Transitional Issues in Moving to a Direct Consumption Tax," *National Tax Journal* 46 (September): 359–76.

Slemrod, Joel (1996), "Which Is the Simplest Tax System of Them All?" in *Economic Effects of Fundamental Tax Reform*, eds. Henry J. Aaron and William G. Gale (Washington, D.C.: Brookings Institution Press), 355–91.

Triest, Robert K. (1996), "Fundamental Tax Reform and Labor Supply," in *Economic Effects of Fundamental Tax Reform*, eds. Henry J. Aaron and William G. Gale (Washington, D.C.: Brookings Institution Press), 247–78.

A NUMERICAL EXAMPLE

It will be useful to consider an example of an economy as described in the main text, which will enable us to quantify the responses to fundamental tax reform. Consider an environment in which households are identical and have preferences described by a utility function of the form

$$\sum_{t=1}^{\infty} \log(c_t)/(1+\rho)^t,$$

where c_t is household consumption at time t, and ρ is the pure rate of time preference. With this utility function, the marginal rate of substitution between current and future consumption $[MRS(c_t, c_{t+1})]$ is simply $(1+\rho)(c_{t+1}/c_t)$. Consistent with Equation 3, the optimality condition $1 + r_t = MRS(c_t, c_{t+1})$ takes the form

$$\frac{c_{t+1}}{c_t} = \frac{1+r_t}{1+\rho},$$

where r_t is the real after-tax interest rate.

We assume that production per household is a simple function of capital per household: $y = k^\theta$. It follows that the marginal product of capital (MP_k) is $\theta k^{\theta-1}$. Investment optimality conditions (Equations 5 and 5′ for an economy with an income tax and an economy with a consumption tax, respectively) become

$$r_t = (1-\tau_p)(\theta k_t^{\theta-1} - \delta)$$

and

$$r_t = \theta k_t^{\theta-1} - \delta,$$

where δ is the depreciation rate for capital and τ_p is the corporate income tax rate. Finally, Equation 7, which governs the evolution of the capital stock, takes the form

$$k_{t+1} - k_t = k_t^\theta - c_t - g_t - \delta k_t,$$

where g_t denotes government purchases at time t.

We assume that a period is a quarter of a year and let $\rho = 0.01$ and $\delta = 0.02$. These parameter values imply that the average annual rate of return on capital is just over 4 percent in steady state, and the annualized depreciation rate is just over 8 percent. The technology parameter, θ, is set equal to 0.35. Government purchases are constant through time.

Figures A, B, and C show the response of the capital stock, consumption, and after-tax rate of return to fundamental tax reform. Initially, the economy is assumed to be in steady state with a corporate income tax rate of 35 percent. (Interest, dividend, and wage taxes may also be in effect, but these are irrelevant to our simulations.) Suddenly, in period 0, the income tax is replaced by a consumption tax. In the figures, the initial steady-state levels of consumption and capital are normalized to unity. As can be seen, convergence is fairly complete after 100 periods, or 25 years. However, consumption does not rise above its initial steady-state level for 40 periods, or 10 years. The after-tax annual interest rate jumps from 4.1 percent to around 6.3 percent before gradually falling back to its old level. As noted in the main text, the pretax interest rate (not shown in the figures) may jump upward or downward when tax reform is implemented, depending upon the relative magnitudes of the tax rates on interest income and corporate profits. Assuming that interest income is initially taxed at a 25 percent annual rate, tax reform would see the pretax interest rate jump upward from 5.4 percent to 6.3 percent, then gradually fall to 4.1 percent.

Figure A
Capital

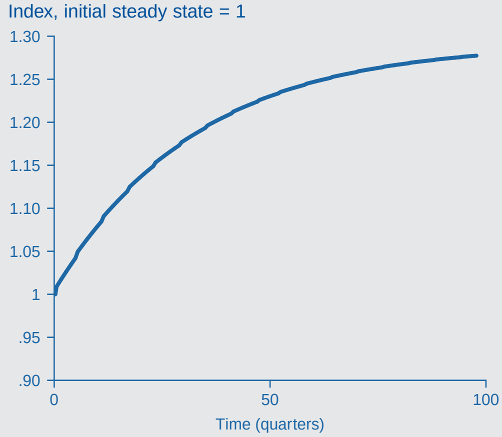


Figure B
Consumption

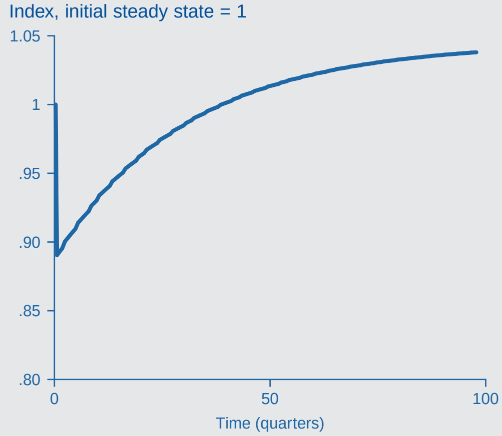


Figure C
After-Tax Interest Rate