Estimation of Dynamic Causal Effects
(SW Chapter 15)

A ***dynamic causal effect*** is the effect on Y of a change in X over time.

For example:

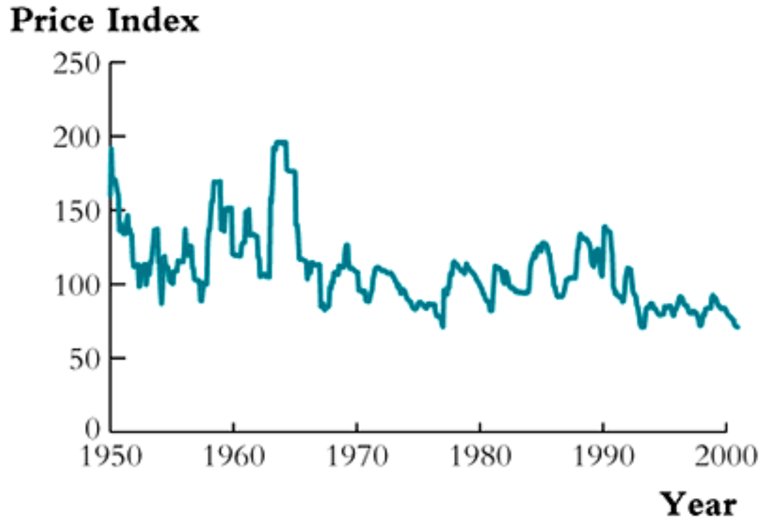
- The effect of an increase in cigarette taxes on cigarette consumption this year, next year, in 5 years;
- The effect of a change in the Fed Funds rate on inflation, this month, in 6 months, and 1 year;
- The effect of a freeze in Florida on the price of orange juice concentrate in 1 month, 2 months, 3 months…

The Orange Juice Data
(SW Section 15.1)

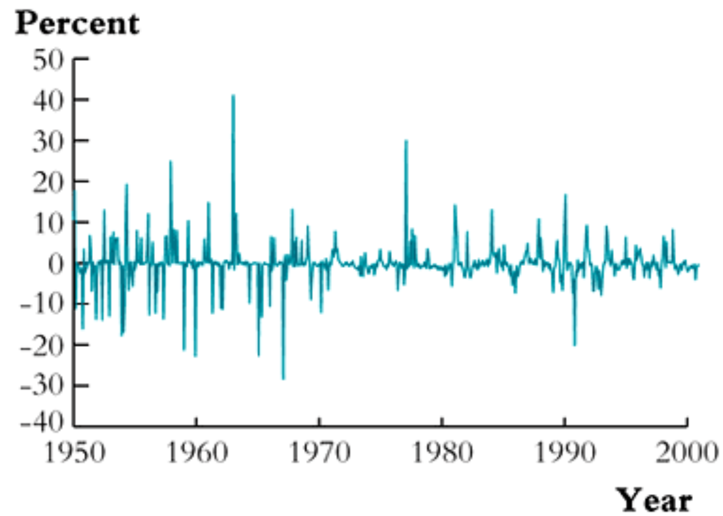
Data

- Monthly, Jan. 1950 – Dec. 2000 ($T = 612$)
- *Price* = price of frozen OJ (a sub-component of the producer price index; US Bureau of Labor Statistics)
- *%ChgP* = percentage change in price at an annual rate, so $\%ChgP_t = 1200\Delta\ln(Price_t)$
- *FDD* = number of freezing degree-days during the month, recorded in Orlando FL
  - Example: If November has 2 days with low temp < 32°, one at 30° and at 25°, then $FDD_{Nov} = 2 + 7 = 9$

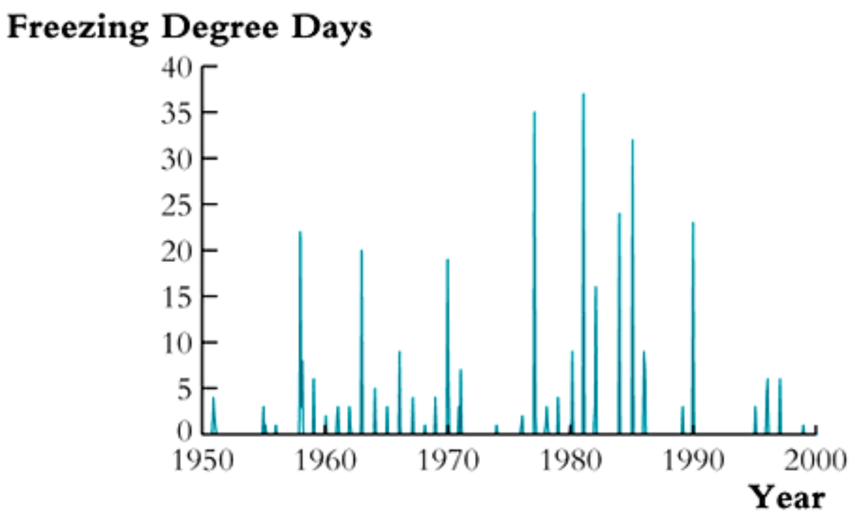
FIGURE 13.1 **Orange Juice Prices and Florida Weather, 1950–2000**



(a) Price Index for Frozen Concentrated Orange Juice



(b) Percent Change in the Price of Frozen Concentrated Orange Juice



(c) Monthly Freezing Degree Days in Orlando, Florida

There have been large month-to-month changes in the price of frozen concentrated orange juice. Many of the large movements coincide with freezing weather in Orlando, home of the orange groves.

Initial OJ regression

$$\widehat{\%ChgP_t} = -.40 + .47FDD_t$$
$$(.22) \quad (.13)$$

- Statistically significant positive relation
- More/deeper freezes, price goes up
- Standard errors: *not the usual – heteroskedasticity and autocorrelation-consistent* (*HAC*) *SE's – more on this later*
- But what is the effect of *FDD* over time?

Dynamic Causal Effects
(SW Section 15.2)

Example: What is the effect of fertilizer on tomato yield?

An ideal randomized controlled experiment

- Fertilize some plots, not others (random assignment)
- Measure yield *over time – over repeated harvests* – to estimate causal effect of fertilizer on:
    - Yield in year 1 of expt
    - Yield in year 2, etc.
- The result (in a large expt) is the causal effect of fertilizer on yield *k* years later.

In time series applications, we can't conduct this ideal randomized controlled experiment:

- We only have one US OJ market …
- We can't randomly assign FDD to different replicates of the US OJ market (?)
- We can't measure the average (across "subjects") outcome at different times – only one "subject"
- So we can't estimate the causal effect at different times using the differences estimator

An alternative thought experiment:

- Randomly give the same subject different treatments (FDD_t) at different times
- Measure the outcome variable ($\%ChgP_t$)
- The "population" of subjects consists of the same subject (OJ market) but at different dates
- If the "different subjects" are drawn from the same distribution – *that is, if* Y_t, X_t *are* ***stationary*** – then the dynamic causal effect can be deduced by OLS regression of Y_t on lagged values of X_t.
- This estimator (regression of Y_t on X_t and lags of X_t) is called the ***distributed lag*** estimator.

Dynamic causal effects and the distributed lag model

The **distributed lag model** is:

$$Y_t = \beta_0 + \beta_1 X_t + \ldots + \beta_{r+1} X_{t-r} + u_t$$

- β_1 = ***impact effect of change in X*** = effect of change in X_t on Y_t, holding past X_t constant
- β_2 = ***1-period dynamic multiplier*** = effect of change in X_{t-1} on Y_t, holding constant X_t, X_{t-2}, X_{t-3},…
- β_3 = ***2-period dynamic multiplier*** (etc.)= effect of change in X_{t-2} on Y_t, holding constant X_t, X_{t-1}, X_{t-3},…
- ***Cumulative dynamic multipliers***
  - *Ex*: the 2-period cumulative dynamic multiplier

$$= \beta_1 + \beta_2 + \beta_3 \text{ (etc.)}$$

Exogeneity in time series regression

Exogeneity (*past and present*)

X is ***exogenous*** if $E(u_t|X_t,X_{t-1},X_{t-2},\ldots) = 0$.

Strict Exogeneity (*past, present, and future*)

X is ***strictly exogenous*** if $E(u_t|\ldots,X_{t+1},X_t,X_{t-1},\ldots) = 0$

- Strict exogeneity implies exogeneity
- For now we suppose that X is exogenous – we'll return (briefly) to the case of strict exogeneity later.
- If X is exogenous then OLS estimates the dynamic causal effect on Y of a change in X. Specifically…

Estimation of Dynamic Causal Effects with Exogenous Regressors (SW Section 15.3)

$$Y_t = \beta_0 + \beta_1 X_t + \ldots + \beta_{r+1} X_{t-r} + u_t$$

The Distributed Lag Model Assumptions

1. $E(u_t|X_t,X_{t-1},X_{t-2},\ldots) = 0$ (X is exogenous)
2. (a) Y and X have stationary distributions;
   (b) (Y_t,X_t) and (Y_{t-j},X_{t-j}) become independent as j gets large
3. Y and X have nonzero finite 8th moments
4. There is no perfect multicollinearity

- Assumptions 1 and 4 are familiar
- Assumption 3 is familiar, except for 8 (not four) finite moments – this has to do with HAC estimators
- Assumption 2 is different – before it was (X_i, Y_i) are i.i.d. – this now becomes more complicated.

2. (a) Y and X have stationary distributions;
    - If so, the coefficients don't change within the sample (internal validity);
    - and the results can be extrapolated outside the sample (external validity).
    - This is the time series counterpart of the "identically distributed" part of i.i.d.

2. (b) (Y_t, X_t) and (Y_{t-j}, X_{t-j}) become independent as j gets large

- Intuitively, this says that we have separate experiments for time periods that are widely separated.
- In cross-sectional data, we assumed that Y and X were i.i.d., a consequence of simple random sampling – this led to the CLT.
- A version of the CLT holds for time series variables that become independent as their temporal separation increases – assumption 2(b) is the time-series counterpart of the "independently distributed" part of i.i.d.

Digression on weak dependence

- Consider the law of large numbers (LLN)
- Let Y_t with $E(Y_t) = \mu$, $var(Y_t) = \sigma^2$ and ACF $= \rho_j$
- The sample mean is $\bar{Y}_T = T^{-1} \sum_{t=1}^{T} Y_t$
- The LLN says that if the data is i.i.d., $\bar{Y}_T \xrightarrow{p} \mu$
- One way to show this is to show that if $var(\bar{Y}_T) \to 0$ as $T \to \infty$
- When Y_t is i.i.d., $var(\bar{Y}_T) = \frac{1}{T^2} \sum_{t=1}^{T} var(Y_t) = \frac{\sigma^2}{T}$
- This indeed goes to zero as $T \to \infty$
- Intuitively, each new observation contributes an independent piece of information

- What happens when $\rho_j = corr(Y_t, Y_{t-j}) \neq 0$?
- Consider $T = 2$

$$var(\bar{Y}_2) = \frac{1}{4}[var(Y_1) + var(Y_2) + 2cov(Y_1, Y_2)]$$

$$= \frac{1}{2}\sigma^2 + \frac{1}{2}\rho_1\sigma^2 = \frac{1}{2}\sigma^2 * f_2, \text{ where } f_2 = (1 + \rho_1)$$

- For general T, it can be shown that

$$var(\bar{Y}_T) = \frac{1}{T}\sigma^2 f_T, \text{ where } f_T = 1 + 2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\rho_j$$

- With i.i.d. data $\rho_j = 0$, so $var(\bar{Y}_T) \to 0$ and LLN holds
- In the other extreme, when $\rho_j = 1$ for all j, then $2\sum_{j=1}^{T-1}\frac{T-j}{T} = T - 1$, so $f_T = T$ and hence $var(\bar{Y}_T) = \sigma^2$ for all T

- In other words, if the data is perfectly autocorrelated, each additional observation provides *no* additional information
- So, the variance of the sample mean does not decrease and the LLN fails to hold
- Assumption 2(b) says that ρ_j goes to zero as j increases, and it does so sufficiently fast so that $\frac{1}{T}\sum_{j=1}^{T-1}\frac{T-j}{T}\rho_j \to 0$ and therefore $var(\bar{Y}_T) \to 0$, as $T \to \infty$
- This assumption is satisfied in stationary AR models, e.g., $\rho_j = \beta_1^j$ where $|\beta_1| < 1$ is the coefficient in an AR(1) model

Under the Distributed Lag Model Assumptions:

- OLS yields consistent estimators of β_1, β_2,…,β_{r+1} (of the dynamic multipliers)
- The sampling distribution of $\hat{\beta}_1$, etc., is normal
- *However*, the formula for the variance of this sampling distribution is not the usual one for i.i.d. data, because u_t may be serially correlated.
- This means that the usual OLS standard errors (usual OxMetrics printout) are wrong unless we are confident that u_t is serially uncorrelated
- Otherwise, we need to use, instead, *SE*s that are robust to autocorrelation as well as to heteroskedasticity…

We need ***H****eteroskedasticity and* ***A****utocorrelation-* ***C****onsistent (HAC) standard errors*

- There are various estimators for HAC standard errors out there
- The oldest and most commonly used is the one developed by Newey and West (1987)
- This is the one reported by OxMetrics
- See SW Section 15.4 for details

Example: OJ and HAC estimators in OxMetrics

- Open file "OJdata.xlsx" and estimate a regression of *ChgP* on *FDD*..

```
EQ( 7) Modelling ChgP by OLS
       The dataset is: /Users/sophocles/Dropbox/Teaching/Oxford/Depar
       The estimation sample is: 1950(1) - 2000(12)

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant            -0.404719     0.1987    -2.04  0.0421   0.0068
fdd                  0.470227    0.05977     7.87  0.0000   0.0921

sigma                 4.83032  RSS                  14232.4929
R^2                 0.0921177  F(1,610) =    61.89 [0.000]**
Adj.R^2             0.0906294  log-likelihood          -1831.23
no. of observations       612  no. of parameters              2
mean(ChgP)          -0.115822  se(ChgP)                  5.0653
```

- The default std errors in OxMetrics are the homoskedastic ones
- To obtain robust std errors, you need to do the following

- Go to "Model\Test", select "Further output", and then "Robust standard errors and t-values"
- This yields the following output

Robust standard errors

	Coefficients	SE	HACSE	HCSE	JHCSE
Constant	-0.40472	0.19868	0.21559	0.18922	0.18997
fdd	0.47023	0.059770	0.13903	0.13840	0.15548

	Coefficients	t-SE	t-HACSE	t-HCSE	t-JHCSE
Constant	-0.40472	-2.0371	-1.8772	-2.1389	-2.1304
fdd	0.47023	7.8672	3.3823	3.3976	3.0243

- HACSE gives Newey-West (1987) HAC standard errors using the default lag truncation $0.75*(612^{1/3}) \approx 7$
- Nonrobust (SE) and White heteroskedasticity robust (HCSE) std errors are also reported for comparison
- In this case the difference is small, but not always so!

- OxMetrics doesn't give much choice over HAC estimators (e.g., no option to alter the lag truncation parameter)
  - Other packages, such as STATA, offer more choices
- That's because the creators of OxMetrics weren't too keen on HAC standard errors
- A well-specified model should adequately explain the dynamics, so it shouldn't have autocorrelated errors (hence, no need to "correct" the std errors)
- However, sometimes we do need them

Example: Dynamic multipliers

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	-0.646905	0.2182	-2.96	0.0032	0.0143
fdd	0.471876	0.05985	7.88	0.0000	0.0933
fdd_1	0.116204	0.05988	1.94	0.0528	0.0062
fdd_2	0.0820362	0.05985	1.37	0.1710	0.0031
fdd_3	0.0686688	0.05986	1.15	0.2518	0.0022
fdd_4	0.0367706	0.05985	0.614	0.5392	0.0006
fdd_5	0.0296313	0.05988	0.495	0.6209	0.0004
fdd_6	0.0592368	0.05985	0.990	0.3227	0.0016

sigma	4.81995	RSS	14032.0989
R^2	0.104901	F(7,604) =	10.11 [0.000]**
Adj.R^2	0.0945271	log-likelihood	-1826.9
no. of observations	612	no. of parameters	8
mean(ChgP)	-0.115822	se(ChgP)	5.0653

Robust standard errors

	Coefficients	SE	HACSE	HCSE
Constant	-0.64691	0.21823	0.23686	0.21321
fdd	0.47188	0.059850	0.14183	0.13995
fdd_1	0.11620	0.059880	0.096723	0.095186
fdd_2	0.082036	0.059854	0.041310	0.045532
fdd_3	0.068669	0.059859	0.048020	0.048836
fdd_4	0.036771	0.059854	0.031066	0.031408
fdd_5	0.029631	0.059880	0.023899	0.024403
fdd_6	0.059237	0.059850	0.045723	0.046000

ADL versus DL model

$$Y_t = \beta_0 + \beta_1 X_t + \ldots + \beta_{r+1} X_{t-r} + u_t$$

- A limitation of the distributed lag model is that it assumes all dynamic multipliers beyond lag r are zero
- Violation of this assumption can lead to omitted variable bias if X_t is autocorrelated
- There are limits to how big one can set r to capture longer period effects
  - You lose observations and degrees of freedom
- An ADL model can do that with few parameters ("parsimoniously" in the lingo)

- Consider the simple ADL(1,1) model

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \gamma_1 Y_{t-1} + u_t$$

- Notice contemporaneous value of X_t on the RHS
- *impact effect of change in* $X = \beta_1$ (effect of change in X_t on Y_t, holding past X_t constant) -- same as for DL model
- ***1-period dynamic multiplier*** $= \boldsymbol{\beta_2 + \gamma_1 \beta_1}$ (effect of change in X_{t-1} on Y_t, holding constant X_t, X_{t-2}, X_{t-3},…)
- Different from DL model:
  - β_2 is direct effect of X_{t-1} on Y_t
  - $\gamma_1 \beta_1$ is indirect effect of X_{t-1} on Y_t through Y_{t-1}

- *2nd-period dynamic multiplier* = $\gamma_1(\beta_2 + \gamma_1\beta_1)$
- This is because

$$\frac{\partial Y_t}{\partial X_{t-2}} = \gamma_1 \frac{\partial Y_{t-1}}{\partial X_{t-2}} = \gamma_1(\beta_2 + \gamma_1\beta_1)$$

- Similarly, we can show that the *j*th-period dynamic multiplier is $\gamma_1^{j-1}(\beta_2 + \gamma_1\beta_1)$
- Thus, in ADL(1,1) model dynamic multipliers are nonzero at all horizons, but decline exponentially to zero

Is there a causal interpretation for the coefficient γ_1?

- If Y_t is consumption, X_t is income, γ_1 can reflect the effect of past consumption on current consumption due to *habits* or *adjustment costs*
- Another example: Y_t is price (set by a firm), X_t is marginal cost of production: it may be costly to adjust prices (think menu costs from your macro lectures)
- γ_1 could reflect expectations
  - $Y_t = \gamma_1 E(Y_{t+1}|Y_{t-1}, \dots) + \cdots$ and $E(Y_{t+1}|Y_{t-1}, \dots) = Y_{t-1}$ (naïve expectations)

  e.g., expectations augmented Phillips curve

Do I need to use HAC *SE*s when I estimate an AR or an ADL model?

NO.

- The problem to which HAC *SE*s are the solution arises when u_t is serially correlated
- If u_t is serially uncorrelated, then OLS *SE*'s are fine
- In AR and ADL models, the errors are serially uncorrelated if you have included enough lags of Y
  - If you include enough lags of Y, then the error term can't be predicted using past Y, or equivalently by past u – so u is serially uncorrelated

Estimation of Dynamic Causal Effects with Strictly Exogenous Regressors (SW Section 15.5)

- X is strictly exogenous if $E(u_t|\ldots,X_{t+1},X_t,X_{t-1},\ldots) = 0$
- If X is strictly exogenous, there are more efficient ways to estimate dynamic causal effects than by a distributed lag regression.
  - Generalized Least Squares (GLS)
- But the condition of strict exogeneity is very strong, and rarely plausible in practice.

Analysis of the OJ Price Data
(SW Section 15.6)

What is the dynamic causal effect (*what are the dynamic multipliers*) of a unit increase in FDD on OJ prices?

$$\%ChgP_t = \beta_0 + \beta_1 FDD_t + \ldots + \beta_{r+1} FDD_{t-r} + u_t$$

- *What r to use*?

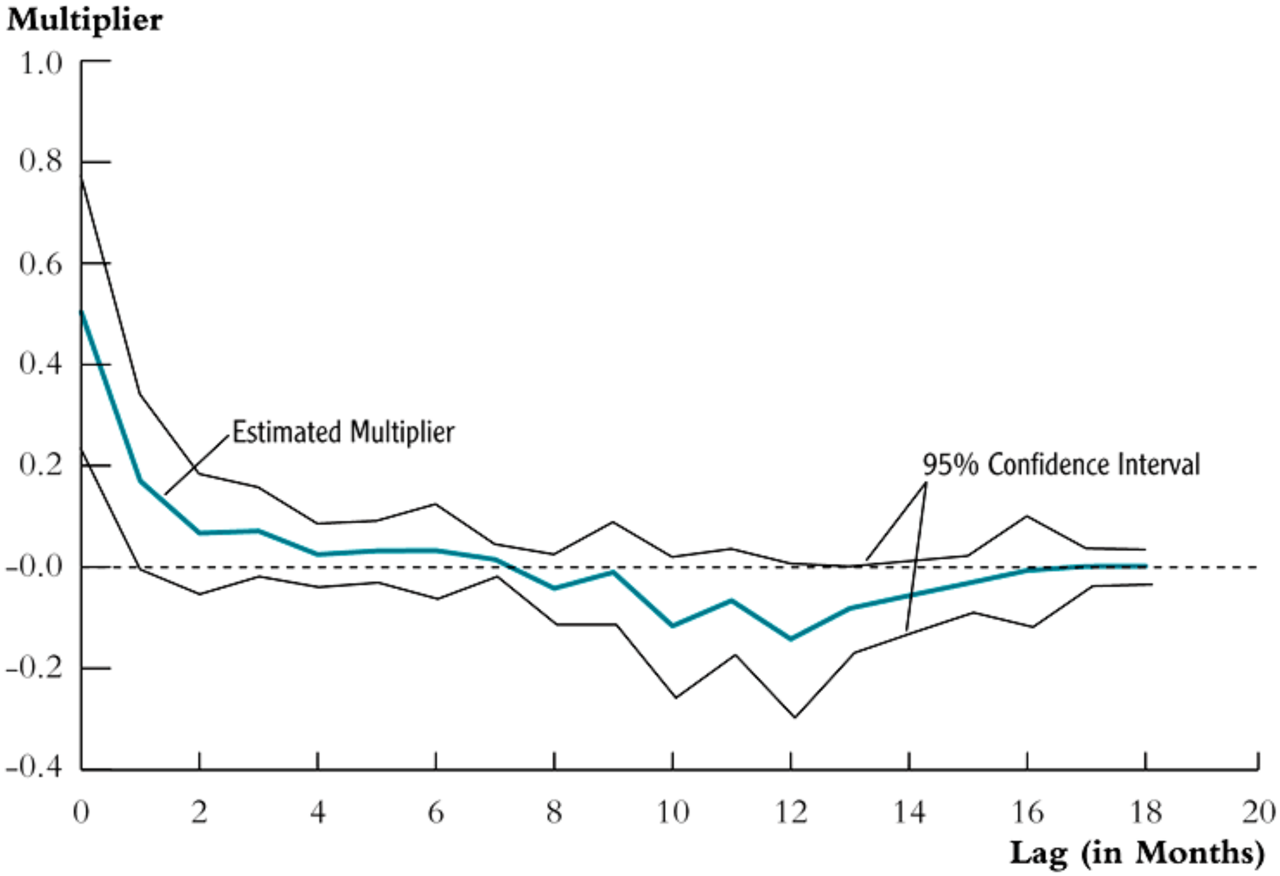
  How about 18? (Goldilocks method)

TABLE 13.1 The Dynamic Effect of a Freezing Degree Day (*FDD*) on the Price of Orange Juice: Selected Estimated Dynamic Multipliers and Cumulative Dynamic Multipliers

Lag number	(1) Dynamic Multipliers	(2) Cumulative Multipliers	(3) Cumulative Multipliers	(4) Cumulative Multipliers
0	0.50 (0.14)	0.50 (0.14)	0.50 (0.14)	0.51 (0.15)
1	0.17 (0.09)	0.67 (0.14)	0.67 (0.13)	0.70 (0.15)
2	0.07 (0.06)	0.74 (0.17)	0.74 (0.16)	0.76 (0.18)
3	0.07 (0.04)	0.81 (0.18)	0.81 (0.18)	0.84 (0.19)
4	0.02 (0.03)	0.84 (0.19)	0.84 (0.19)	0.87 (0.20)
5	0.03 (0.03)	0.87 (0.19)	0.87 (0.19)	0.89 (0.20)
6 ⋮	0.03 (0.05)	0.90 (0.20)	0.90 (0.21)	0.91 (0.21)
12 ⋮	−0.14 (0.08)	0.54 (0.27)	0.54 (0.28)	0.54 (0.28)
18	0.00 (0.02)	0.37 (0.30)	0.37 (0.31)	0.37 (0.30)
Monthly indicators?	No	No	No	Yes $F = 1.01$ $(p = 0.43)$
HAC standard error truncation parameter (m)	7	7	14	7

All regressions were estimated by OLS using monthly data (described in Appendix 13.1) from January 1950 to December 2000, for a total of $T = 612$ monthly observations. The dependent variable is the monthly percentage change in the price of orange juice ($\%ChgP_t$). Regression (1) is the distributed lag regression with the monthly number of freezing degree days and eighteen of its lagged values, that is, $FDD_t, FDD_{t-1}, \ldots, FDD_{t-18}$, and the reported coefficients are the OLS estimates of the dynamic multipliers. The cumulative multipliers are the cumulative sum of estimated dynamic multipliers. All regressions include an intercept, which is not reported. Newey-West HAC standard errors, computed using the truncation number given in the final row, are reported in parentheses.

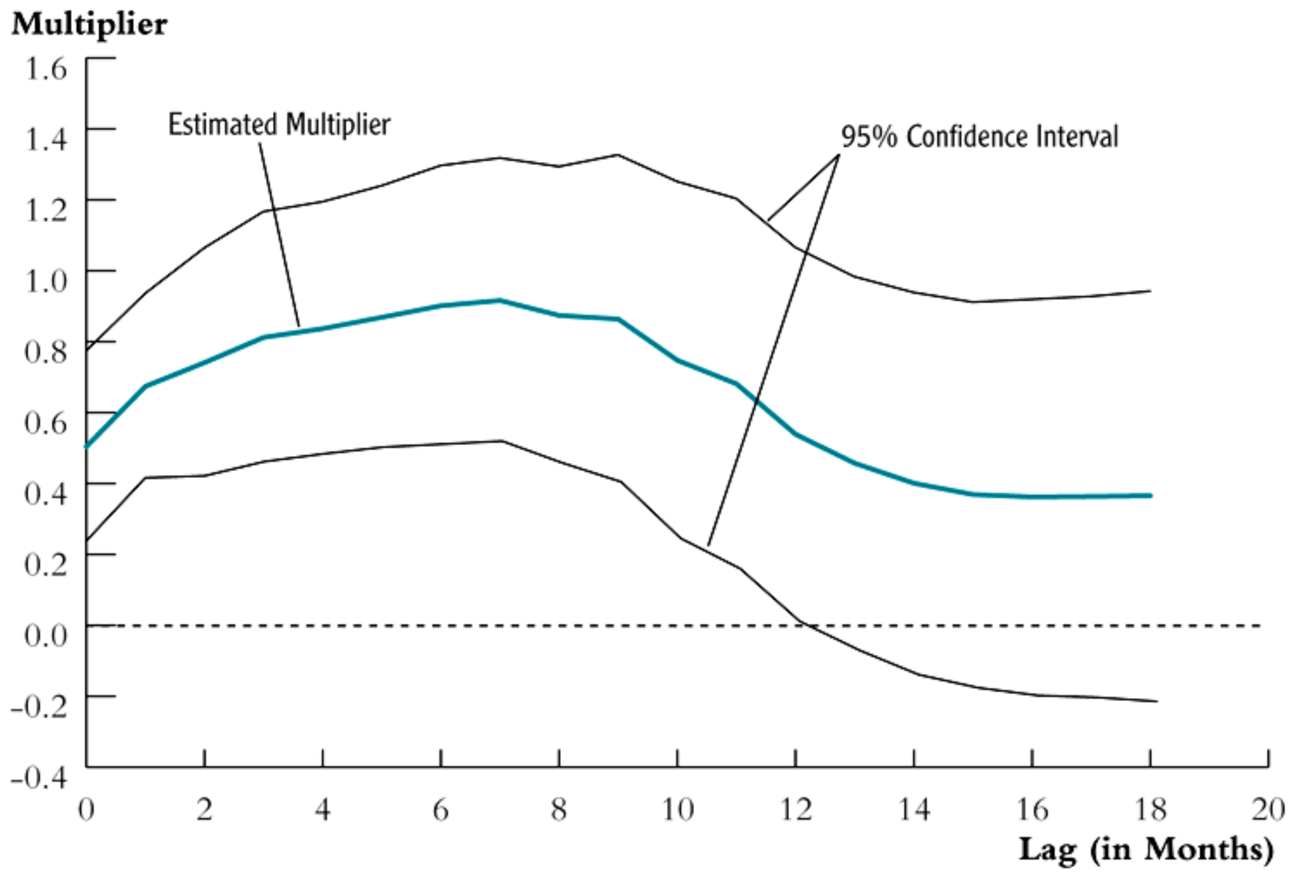
FIGURE 13.2 The Dynamic Effect of a Freezing Degree Day (*FDD*) on the Price of Orange Juice



(a) Estimated Dynamic Multipliers and 95% Confidence Interval

The estimated dynamic multipliers show that a freeze leads to an immediate increase in prices. Future price rises are much smaller than the initial impact. The cumulative multiplier shows that freezes have a persistent effect on the level of orange juice prices, with prices peaking seven months after the freeze.

FIGURE 13.2 **The Dynamic Effect of a Freezing Degree Day (*FDD*) on the Price of Orange Juice**

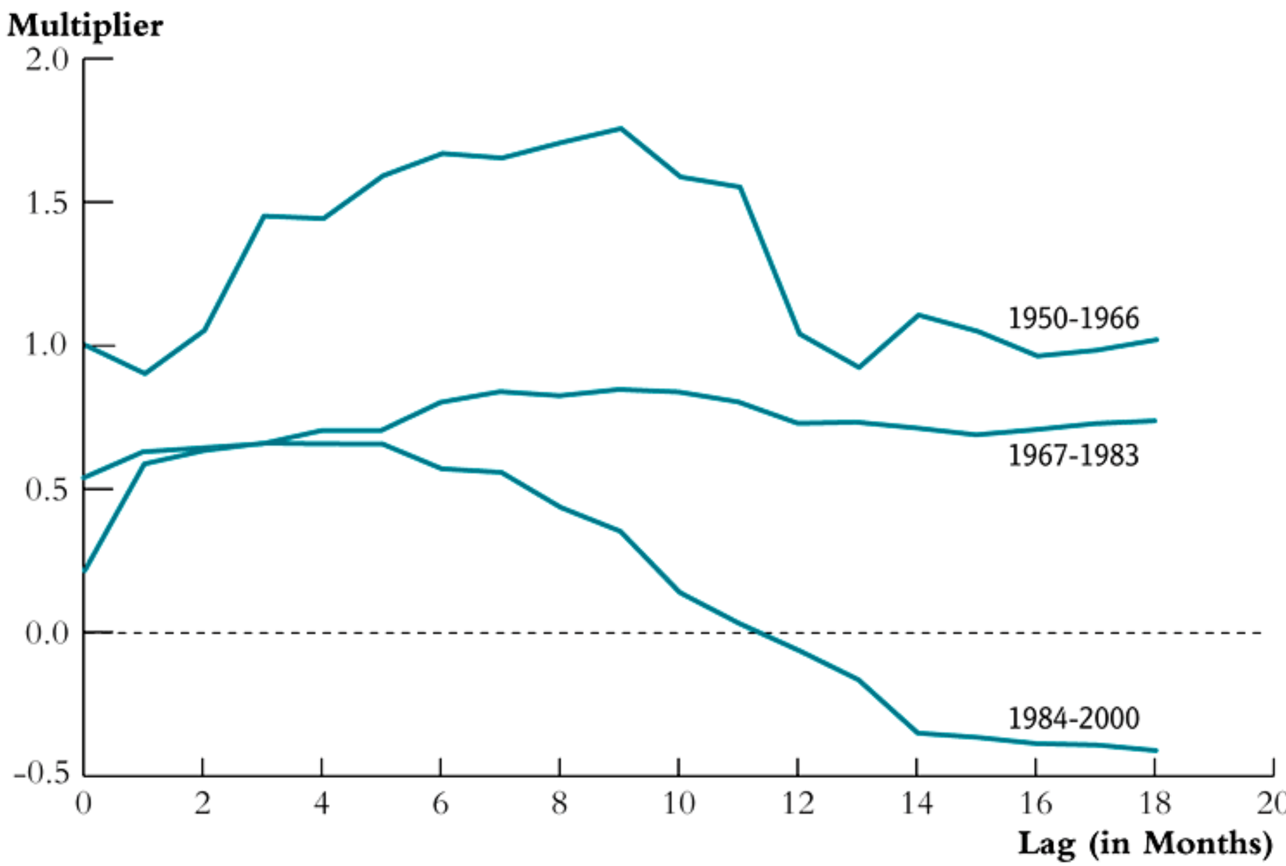


(b) Estimated Cumulative Dynamic Multipliers and 95% Confidence Interval

The estimated dynamic multipliers show that a freeze leads to an immediate increase in prices. Future price rises are much smaller than the initial impact. The cumulative multiplier shows that freezes have a persistent effect on the level of orange juice prices, with prices peaking seven months after the freeze.

FIGURE 13.3 **Estimated Cumulative Dynamic Multipliers from Different Sample Periods**

The dynamic effect on orange juice prices of freezes changed significantly over the second half of the twentieth century. A freeze had a larger impact on prices during 1950 to 1966 than later, and the effect of a freeze was less persistent during 1984–2000 than earlier.



These dynamic multipliers were estimated using a distributed lag model. Should we attempt to obtain more efficient estimates using GLS?

- Is *FDD* strictly exogenous in the distributed lag regression?

$$\%ChgP_t = \beta_0 + \beta_1 FDD_t + \ldots + \beta_{r+1} FDD_{t-r} + u_t$$

- OJ commodity traders can't change the weather.
- So this implies that corr(u_t,FDD_{t+1}) = 0, right?

When Can You Estimate Dynamic Causal Effects? That is, When is Exogeneity Plausible? (SW Section 15.7)

In the following examples,

- is X exogenous?
- is X strictly exogenous?

Examples:

1. Y = OJ prices, X = FDD in Orlando
2. Y = Australian exports, X = US GDP (effect of US income on demand for Australian exports)

Examples, ctd.

3. Y = EU exports, X = US GDP (effect of US income on demand for EU exports)
4. Y = US rate of inflation, X = percentage change in world oil prices (as set by OPEC) (effect of OPEC oil price increase on inflation)
5. Y = GDP growth, X =Federal Funds rate (the effect of monetary policy on output growth)
6. Y = change in the rate of inflation, X = unemployment rate on inflation (the Phillips curve)

Exogeneity, ctd.

- You must evaluate exogeneity and strict exogeneity on a case by case basis
- Exogeneity is often not plausible in time series data because of simultaneous causality
- Strict exogeneity is rarely plausible in time series data because of feedback

Estimation of Dynamic Causal Effects: Summary (SW Section 15.8)

- Dynamic causal effects are measurable in theory using a randomized controlled experiment with repeated measurements over time.
- When X is exogenous, you can estimate dynamic causal effects using a distributed lag regression
- If u is serially correlated, conventional OLS *SE*s are incorrect; you must use HAC *SE*s
- To decide whether X is exogenous, think hard!