Estimation of Dynamic Causal Effects

606EC Econometria - Mod. II

2023-2023

Martin Magris

Overview

Introduction

Introduction I

This chapter takes up the problem of estimating the effect on Y now and in the future of a change in X - that is, the dynamic causal effect on Y of a change in X.

Example: What is the effect on the path of orange juice prices over time of a freezing spell in Florida?

...The starting point for modeling and estimating dynamic causal effects is the so-called distributed lag regression model, in which Y_t is expressed as a function of current and past values of X_t.

Introduction II

Three approaches:

1. One way to estimate dynamic causal effects is to estimate the coefficients of the distributed lag regression model using ordinary least squares (OLS).
2. A second way to estimate dynamic causal effects, is to model the serial correlation in the error term as an autoregression and then use this autoregressive model to derive an autoregressive distributed lag (ADL) model
3. Alternatively, the coefficients of the original distributed lag model can be estimated by generalized least squares (GLS).

Introduction III

Both the ADL and the GLS methods, however, require a stronger version of exogeneity than we have used so far: *strict exogeneity*, under which the regression errors have a conditional mean of 0 given past, present, and *future* values of X.

An Initial Taste of the Orange Juice Data I

- Orlando, the historical center of Florida's orange-growing region, is normally sunny and warm. But now and then there is a cold snap, and if temperatures drop below freezing for too long, the trees drop many of their oranges
- If the cold snap is severe, the trees freeze. Following a freeze, the supply of orange juice concentrate falls, and its price rises.
- The timing of the price increases is rather complicated, however

An Initial Taste of the Orange Juice Data II

Orange juice concentrate is a "durable", or storable, commodity:

$\rightarrow$ The price of orange juice concentrate depends not only on current supply but also on expectations of future supply.

A freeze today means that future supplies of concentrate will be low, but because concentrate currently in storage can be used to meet either current or future demand, the price of existing concentrate rises today...

Question: how much does the price of concentrate rise when there is a freeze

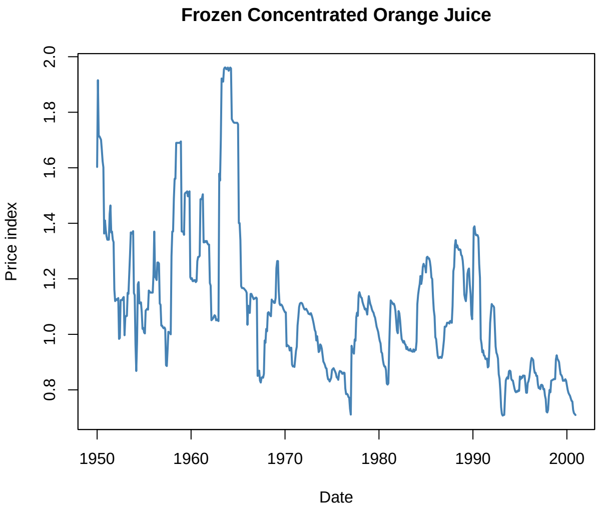
An Initial Taste of the Orange Juice Data III

Data:

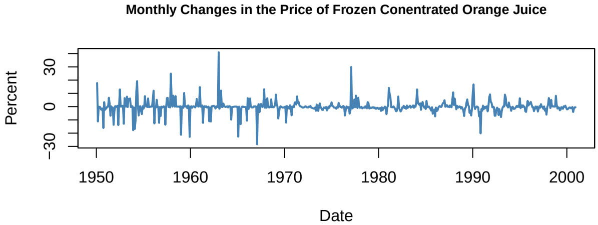
- Monthly data on the price of frozen orange juice concentrate, its monthly per centage change, and temperatures in the orange-growing region of Florida from January 1950 to December 2000
- This price was deflated by the overall producer price index for finished goods to eliminate the effects of overall price inflation
- The temperature data are the number of freezing-degree days, calculated as the sum of the number of degrees Fahrenheit that the minimum temperature falls below freezing on a given day over all days in the month.

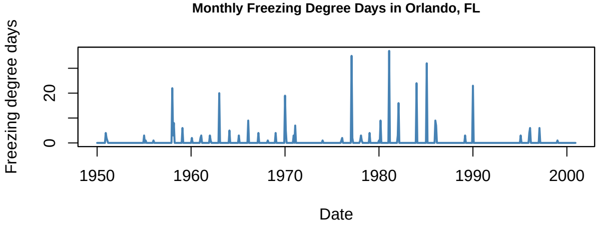
  Example: In Nov. 1950 the airport temperature dropped below 25th (31°F) and on the 29th (29°F), for a total of 4 freezing degree days $[(32-31)+(32-29)]=4$.
- The price of orange juice concentrate has large swings, some of which appear to be associated with cold weather in Florida.

An Initial Taste of the Orange Juice Data IV



An Initial Taste of the Orange Juice Data V





An Initial Taste of the Orange Juice Data VI

Use a regression to estimate the amount by which orange juice prices rise when the weather turns cold

- The dependent variable is the percentage change in the price over that month ($\% ChgOJC_t = 100 \times \Delta \ln OJC_t$)
- The regressor is the number of freezing degree days during that month FDD_t
- This regression is estimated using monthly data from January 1950 to December 2000, $T = 612$ observations

$$\% \widehat{ChgOJC}_t = \underset{(0.19)}{-0.42} + \underset{(0.13)}{0.47}\, FDD_t$$

(the standard errors are HAC standard errors)

An Initial Taste of the Orange Juice Data VII

The estimated coefficient on FDD_t has the following interpretation:

- an additional freezing degree day in month $t$leads to a price increase of 0.47 percentage points in the same month.
- In a month with 4 freezing degree days, such as November 1950, the price of orange juice concentrate is estimated to have increased by $4 \times 0.47\% = 1.88\%$ relative to a month with no days below freezing.

To consider the effects of cold snaps on the orange juice price over the subsequent periods, we include lagged values of FDD_t in our model, which leads to a distributed lag regression model. We estimate a specification using the contemporaneous and six lagged values of FDD_t as regressors.

An Initial Taste of the Orange Juice Data VIII

$$
\begin{aligned}
\%\widehat{ChgOJC_t} = & -\underset{(0.21)}{0.69} + \underset{(0.14)}{0.47}\, FDD_t \\
& + \underset{(0.08)}{0.15}\, FDD_{t-1} + \underset{(0.06)}{0.06}\, FDD_{t-2} + \underset{(0.05)}{0.07}\, FDD_{t-3} \\
& + \underset{(0.03)}{0.04}\, FDD_{t-4} + \underset{(0.03)}{0.05}\, FDD_{t-5} + \underset{(0.05)}{0.05}\, FDD_{t-6}, \qquad (1)
\end{aligned}
$$

Where:

- where the coefficient on FDD_{t-1} estimates the price increase in period t caused by an additional freezing degree day in the preceding month, the coefficient on FDD_{t-2} estimates the effect of an additional freezing degree day two months ago, and so on.

An Initial Taste of the Orange Juice Data IX

- the coefficients in (1) can be interpreted as price changes in current and future periods due to a unit increase a past month's freezing degree days.
- the 4 freezing degree days in November 1950 are estimated to have increased orange juice prices by 1.88% during November 1950, by an additional $4 \times 0.15\% = 0.6\%$ in December 1950, by an additional $4 \times 0.07 = 0.28\%$ in January 1951, and so forth.

Dynamic Causal Effects

Dynamic effects and the distributed lag model I

What, precisely, is meant by a dynamic causal effect?

Because dynamic effects necessarily occur over time, the econometric model used to estimate dynamic causal effects needs to incorporate lags.

To do so, Y_t can be expressed as a distributed lag of current and r past values of X_t:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \cdots + \beta_{r+1} X_{t-r} + u_t, \qquad (2)$$

where u_t is an error term that includes the measurement error in Y_t and the effect of omitted determinants of Y_t.

The model in eq.(2) is called the **distributed lag** model relating X_t, and r of its lags, to Y_t.

Dynamic effects and the distributed lag model II

- The coefficient on the contemporaneous value of X_t, β_1, is the contemporaneous or immediate effect of a unit change in X_t on Y_t.
- The coefficient on X_{t-1}, β_2, is the effect on Y_t of a unit change in X_{t-1} or, equivalently, the effect on Y_{t+1} of a unit change in X_t:
$\rightarrow$ β_2 is the effect of a unit change in X on Y one period later.
- The coefficient on $X_{t-h}$is the effect of a unit change in X on Y after h periods.
- The dynamic causal effect is the effect of a change in X_t on $Y_t, Y_{t+1}, Y_{t+2}, \ldots$

$\rightarrow$ in the context of the distributed lag model in eq. (2) the dynamic causal effect is the sequence of coefficients $\beta_1, \beta_2, \ldots, \beta_{r+1}$

Implications for empirical time series analysis I

This formulation of dynamic causal effects in time series data implies that for the empirical attempts to measure the dynamic causal effect with observational time series data:

i The dynamic causal effect should not change over the sample on which we have data (is implied by the data being jointly stationary)

ii X must be uncorrelated with the error term

Two Types of Exogeneity I

We have defined an *exogenous* variable as a variable that is uncorrelated with the regression error term and an *endogenous* variable as a variable that is correlated with the error term.

This terminology traces to models with multiple equations, in which an endogenous variable is determined within the model, while an exogenous variable is determined outside the model.

if we are to estimate dynamic causal effects using the distributed lag model in eq. (2), the regressors must be uncorrelated with the error term. Thus regressors must be exogenous, but at the same time are determined within the model...

$\rightarrow$ we need to refine the definitions of exogeneity.

Two Types of Exogeneity II

The first concept of exogeneity is that the error term has a conditional mean of 0 given current and all past values of X_t:

$$\mathbb{E}(u_t|X_t, X_{t-1}, X_{t-2}, \ldots) = 0.$$

This modifies the standard conditional mean assumption for multiple regression with cross-sectional data, which requires only that u_t have a conditional mean of 0 given the included regressors-that is, $\mathbb{E}(u_t|X_t, X_{t-1}, X_{t-2}, \ldots, X_{t-r}) = 0$.

We can refer to the assumption that $\mathbb{E}(u_t|X_t, X_{t-1}, X_{t-2}, \ldots) = 0$ as **past and present exogeneity**, or simply **exogeneity**.

Two Types of Exogeneity III

The second concept of exogeneity is that the error term has mean 0 given all *past, present, and future values* of X_t,

$$\mathbb{E}(u_t| \ldots, X_{t+2}, X_{t+1} X_t, X_{t-1}, X_{t-2}, \ldots) = 0.$$

This is called **strict exogeneity** or **past, present, and future exogeneity**.

The reason for introducing the concept of strict exogeneity is that, when X is *strictly* exogenous, there are more efficient estimators of dynamic causal effects than the OLS estimators of the coefficients of the distributed lag regression in eq. (2).

Two Types of Exogeneity IV

Thus strict exogeneity implies exogeneity but not the reverse.

- If X is (past and present) exogenous, then u_t is uncorrelated with current and past values of X_t
- If X is strictly exogenous, then in addition u_t is uncorrelated with future values of X_t.
- If a change in Y_t causes future values of X_t to change, then X_t is not strictly exogenous even though it might be (past and present) exogenous.

Two Types of Exogeneity V

Example: consider the hypothetical multiyear tomato/fertilizer experiment following Equation (16.3). Because the fertilizer is randomly applied in the hypothetical experiment, it is exogenous. Because tomato yield today does not depend on the amount of fertilizer applied in the future, the fertilizer time series is also strictly exogenous.

Two Types of Exogeneity VI

Example: consider the orange juice price example, in which Y_t is the monthly percentage change in orange juice prices and X_t is the number of freezing degree days in that month.

- we can think of the weather, the number of freezing degree days, as if it were randomly assigned in the sense that the weather is outside human control
- If the effect of FDD is linear and if it has no effect on prices after r months, then it follows that the weather is exogenous

But is the weather *strictly* exogenous?

Two Types of Exogeneity VII

If the conditional mean of u_t given future *FDD* is nonzero, then *FDD* is not strictly exogenous...

- if orange juice market participants use forecasts of *FDD* when they decide how much they will buy or sell at a given price, then orange juice prices, and thus the error term u_t, could incorporate information about future *FDD* that would make u_t a useful predictor of *FDD*
- This means that u_t will be correlated with future values of *FDD*
- becauseu_t includes forecasts of future Florida weather, *FDD* would be (past and present) exogenous but not strictly exogenous

Two Types of Exogeneity VIII

To sum up, while tomato plants *are not* affected by future fertilization, orange juice market participants *are* influenced by forecasts of future Florida weather.

Two Types of Exogeneity IX

The Distributed Lag Model and Exogeneity

In the distributed lag model

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \cdots + \beta_{r+1} X_{t-r} + u_t, \qquad (16.4)$$

there are two different types of exogeneity—that is, two different exogeneity conditions:

- Past and present exogeneity (exogeneity):

$$E(u_t \mid X_t, X_{t-1}, X_{t-2}, \ldots) = 0; \text{ and} \qquad (16.5)$$

- Past, present, and future exogeneity (strict exogeneity):

$$E(u_t \mid \ldots, X_{t+2}, X_{t+1}, X_t, X_{t-1}, X_{t-2}, \ldots) = 0. \qquad (16.6)$$

If X is strictly exogenous, it is exogenous, but exogeneity does not imply strict exogeneity.

Estimation of Dynamic Causal Effects with Exogenous Regressors

Estimation of Dynamic Causal Effects with Exogenous Regressors

If X is exogenous, then its dynamic causal effect on Y can be estimated by OLS estimation of the distributed lag regression in eq.(2).

This section summarizes the conditions under which these OLS estimators lead to valid statistical inferences and introduces dynamic multipliers and cumulative dynamic multipliers.

The Distributed Lag Model Assumptions I

The four assumptions of the distributed lag regression model are similar to the four assumptions for the cross-sectional multiple regression model, but they have been modified for time series data.

The Distributed Lag Model Assumptions

The distributed lag model is given in Key Concept 16.1 [Equation (16.4)], where $\beta_1, \beta_2, \ldots, \beta_{r+1}$ are dynamic causal effects and

1. X is exogenous; that is, $E(u_t \mid X_t, X_{t-1}, X_{t-2}, \ldots) = 0$;
2. (a) The random variables Y_t and X_t have a stationary distribution, and (b) (Y_t, X_t) and (Y_{t-j}, X_{t-j}) become independent as j gets large;
3. Large outliers are unlikely: Y_t and X_t have more than eight nonzero finite moments; and
4. There is no perfect multicollinearity.

The Distributed Lag Model Assumptions II

1 The first assumption is that X is exogenous, which extends the 0 conditional mean assumption for cross-sectional data to include all lagged values of X

2a The variables have a stationary distribution,

2b they become independently distributed when the amount of time separating them becomes large

(These assumptions are the same as the corresponding assumption for the ADL model)

The Distributed Lag Model Assumptions III

3. The third assumption is that large outliers are unlikely, by assuming that the variables have more than *eight* nonzero finite moments.

   This is stronger than the assumption of four finite moments, used in the mathematics behind the HAC variance estimator.

4. same as that in the cross-sectional multiple regression model

Note: The distributed lag model extends directly to multiple X's. The additional X's and their lags are simply included as regressors in the distributed lag regression and the assumptions above are modified to include these additional regressors.

Autocorrelated u_t, Standard Errors, and Inference I

In the distributed lag regression model, the error term u_t *can* be autocorrelated: u_t can be correlated with its lagged values.

This autocorrelation arises because, in time series data, the omitted factors included in u_t can themselves be serially correlated.

Example: The demand for orange juice also depends on income, so one factor that influences the price of orange juice is income (of potential orange juice consumers). Then aggregate income is an omitted variable in the distributed lag regression of orange juice price changes against freezing degree days. Aggregate income is serially correlated: Income tends to fall in recessions and rise in expansions. Thus income is serially correlated, and because it is part of the error term, u_t will be serially correlated.

Autocorrelated u_t, Standard Errors, and Inference II

The autocorrelation of u_t does *not* affect the consistency of OLS, nor does it introduce bias.

- If the errors are autocorrelated, the usual OLS standard errors are *inconsistent*.
- When the errors are serially correlated, standard errors predicated on independently and i.i.d. errors are "wrong" in the sense that they result in misleading statistical inferences.
- A solution to this problem is to use HAC standard errors.

Dynamic Multipliers and Cumulative Dyn. Multipliers I

Another name for the dynamic causal effect is the **dynamic multiplier**.

The cumulative dynamic multipliers are the *cumulative* causal effects, up to a given lag; thus the cumulative dynamic multipliers measure the *cumulative effect* on Y of a change in X.

Dynamic Multipliers and Cumulative Dyn. Multipliers II

Definition (Dynamic multipliers)

The effect of a unit change in X on Y after h periods, which is β_{h+1} in eq. (2), is called the h-period **dynamic multiplier**.

$\rightarrow$ Thus the dynamic multipliers relating X to Y are the coefficients on X_t and its lags in (2).

The zero-period (or contemporaneous) dynamic multiplier, or impact effect, is β_1, the effect on Y of a change in X in the *same* period.

Dynamic Multipliers and Cumulative Dyn. Multipliers III

Definition (Cumulative dynamic multipliers)

The h-period cumulative dynamic multiplier is the cumulative effect of a unit change in X on Y over the next h periods. Thus the cumulative dynamic multipliers are the cumulative sum of the dynamic multipliers.

In terms of the coefficients of the distributed lag regression in eq. (2), the zero-period cumulative multiplier is β_1, the one-period cumulative multiplier is $\beta_1 + \beta_2$, and the h-period cumulative dynamic multiplier is $\beta_1 + \beta_2 + \cdots + \beta_{h+1}$.

The sum of all the individual dynamic multipliers, $\beta_1 + \beta_2 + \cdots + \beta_{r+1}$ is the cumulative long-run effect on Y of a change in X and is called the **long-run cumulative dynamic multiplier**.

Dynamic Multipliers and Cumulative Dyn. Multipliers IV

Example: consider eq. (1),

- The immediate effect of an additional freezing degree day is that the price of orange juice concentrate rises by 0.47%
- The cumulative effect of a price change over the next month is the sum of the impact effect and the dynamic effect one month ahead, $0.47\% + 0.15\% = 0.62\%$
- The cumulative dynamic multiplier over two months is $0.47\% + 0.15\% + 0.06\% = 0.68\%$.

Dynamic Multipliers and Cumulative Dyn. Multipliers V

The cumulative dynamic multipliers can be estimated directly using a modification of the distributed lag regression in eq.(2).
The modified regression is,

$$Y_t = \delta_0 + \delta_1 \Delta X_t + \delta_2 \Delta X_{t-1} + \cdots + \delta_r \Delta X_{t-r+1} + \delta_{r+1} X_{t-r} + u_t \quad (3)$$

- The cumulative dynamic multipliers of the distributed lag model in eq.(2) are the coefficients $\delta_1, \delta_2, \ldots, \delta_r, \delta_{r+1}$ in the modified regression.
- Eq.(3) can be directly estimated using OLS which makes it convenient to compute their HAC standard errors.

Dynamic Multipliers and Cumulative Dyn. Multipliers VI

- This can be shown that the population regressions in eq.(3) and eq.(2) are equivalent, where $\delta_0 = \beta_1$, $\delta_2 = \beta_1 + \beta_2$, $\delta_3 = \beta_1 + \beta_2 + \beta_2$, $\delta_4 = \dots$.
- $\delta_{r+1} = \beta_1 + \beta_2 + \cdots + \beta_{r+1}$ is the **long-run cumulative multiplier**.
- The OLS estimators of the coefficients in eq.(3) are the same as the corresponding cumulative sum of the OLS estimators in eq.(2). For example, $\hat{\delta}_2 = \hat{\beta}_1 + \hat{\beta}_2$.
- The HAC standard errors of the coefficients in eq.(3) are the HAC standard errors of the cumulative dynamic multipliers.

Heteroskedasticity- and Autocorrelation- Consistent Standard Errors

Distribution of the OLS Estim. with Autocorr. Errors I

The error term u_t in the distributed lag model (2) may be serially correlated due to serially correlated determinants of Y_t that are not included as regressors.
When these factors are not correlated with the regressors included in the model, serially correlated errors do not violate the assumption of exogeneity such that the OLS estimator remains unbiased and consistent.

However, autocorrelated standard errors render the usual homoskedasticity-only and heteroskedasticity-robust standard errors invalid and may cause misleading inferences. HAC errors are a remedy.

Distribution of the OLS Estim. with Autocorr. Errors II

HAC Standard Errors

The problem: The error term u_t in the distributed lag regression model in Key Concept 16.1 can be serially correlated. If so, the OLS coefficient estimators are consistent, but, in general, the usual OLS standard errors are not, resulting in misleading hypothesis tests and confidence intervals.

The solution: Standard errors should be computed using a HAC estimator of the variance. The HAC estimator involves estimates of $m - 1$ autocorrelations as well as the variance; in the case of a single regressor, the relevant formulas are given in Equations (16.15) and (16.16).

In practice, using HAC standard errors entails choosing the truncation parameter m. To do so, use the formula in Equation (16.17) as a benchmark and then increase or decrease m, depending on whether your regressors and errors have high or low serial correlation.

Distribution of the OLS Estim. with Autocorr. Errors III

It can be shown that, with autocorrelated errors **[PROOF]**:

$$\mathbb{V}\left(\hat{\beta}_1\right) = \left[\frac{1}{T}\frac{\sigma_v^2}{\left(\sigma_X^2\right)^2}\right] f_T \tag{4}$$

where T is the sample size, $v_t = (X_t - \mu_X)u_t$, $\sigma_v^2 = \mathbb{V}(v_t)$, and

$$f_T = 1 + 2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\rho_j, \qquad \rho_j = \mathbb{C}\text{orr}(v_t, v_{t-j}). \tag{5}$$

If the factor f_T, were known, then the variance of $\hat{\beta}_1$ could be estimated by multiplying the usual cross-sectional estimator of the variance by f_T. However, the unknown autocorrelations of v_t must be estimated.

Proof: variance of $\hat{\beta}_1$ with correlated errors I

To show: that the variance of $\hat{\beta}_1$ is that of eq.(4), with f_T given in eq.(5).

The main idea is to show that $\mathbb{V}\left(\hat{\beta}_1\right)$ can be written as the product of two terms:

i the $\mathbb{V}\left(\hat{\beta}_1\right)$ as if u_t is not serially correlated (term into the brackets in eq.(4))

ii a correction factor for the autocorrelation in u_t (term f_t in eq.(4))

Recall, that

$$\hat{\beta}_1 \overset{4.28}{=} \beta_1 + \frac{\frac{1}{T}\sum_t \left(X_t - \bar{X}\right) u_t}{\frac{1}{T}\sum_t \left(X_t - \bar{X}\right)^2},$$

and that, in large samples,

$$\bar{X} \underset{KC.3.1}{\overset{p}{\rightarrow}} \mu_X, \qquad \frac{1}{T}\sum_t \left(X_t - \bar{X}\right)^2 \underset{3.8}{\overset{p}{\rightarrow}} \sigma_X^2.$$

Proof: variance of $\hat{\beta}_1$ with correlated errors II

Therefore

$$\hat{\beta}_1 - \beta_1 \approx \frac{\frac{1}{T}\sum_t (X_t - \mu_X) u_t}{\sigma_X^2} = \frac{\frac{1}{T}\sum_t v_t}{\sigma_X^2} = \frac{\bar{v}}{\sigma_X^2}$$

with

$$v_t = (X_t - \mu_X) u_t, \qquad \bar{v} = \frac{1}{T}\sum v_t.$$

Since β_1 and σ_X^2 are non-random, (they are unknown but nevertheless numbers),

$$\mathbb{V}\left(\hat{\beta}_1 - \beta_1\right) \overset{2.13}{=} \mathbb{V}\left(\hat{\beta}_1\right) \overset{2.13}{=} \frac{\mathbb{V}(\bar{v})}{\left(\sigma_X^2\right)^2}. \tag{6}$$

If v_t are i.i.d., as for cross-sectional data,

$$\mathbb{V}(\bar{v}) \overset{3.9}{=} \frac{\mathbb{V}(v_t)}{T} = \frac{\sigma_v^2}{T},$$

Proof: variance of $\hat{\beta}_1$ with correlated errors III

so that eq.(6) rewrites as:

$$\mathbb{V}\left(\hat{\beta}_1\right) \overset{KC.4.4}{=} \frac{1}{T}\frac{\mathbb{V}[(X_t - \mu_x)u_t]}{[\mathbb{V}(X_t)]^2} \overset{KC.4.3}{=} \frac{1}{T}\frac{\sigma_v^2}{\left(\sigma_X^2\right)^2}. \tag{7}$$

$\rightarrow$ this is the expression of the variance of $\hat{\beta}_1$ in the "ideal" case where v_t are i.i.d.

However, if u_t and X_t are not independently distributed over time, then v_t (defined upon u_t and X_t) will be serially autocorrelated, so that

$$\mathbb{V}(\bar{v}) \neq \frac{1}{T}\mathbb{V}(v_t),$$

and eq.(7) does not hold true.

Proof: variance of $\hat{\beta}_1$ with correlated errors IV

In fact,

$$\begin{aligned}
\mathbb{V}(\bar{v}) &= \mathbb{V}\left[\frac{1}{T}(v_1 + v_2 + \cdots + v_T)\right] = \frac{1}{T^2}\mathbb{V}(v_1 + v_2 + \cdots + v_T) \\
&= \frac{1}{T^2}[\\
&\quad\ \mathbb{V}(v_1) + \mathbb{C}\mathrm{ov}(v_1, v_2) + \mathbb{C}\mathrm{ov}(v_1, v_3) + \cdots + \mathbb{C}\mathrm{ov}(v_1, v_T) \\
&\quad + \mathbb{V}(v_2) + \mathbb{C}\mathrm{ov}(v_1, v_2) + \mathbb{C}\mathrm{ov}(v_2, v_3) + \cdots + \mathbb{C}\mathrm{ov}(v_2, v_T) \\
&\quad + \ldots \\
&\quad + \mathbb{V}(v_T) + \mathbb{C}\mathrm{ov}(v_1, v_T) + \mathbb{C}\mathrm{ov}(v_2, v_T) + \cdots + \mathbb{C}\mathrm{ov}(v_{T-1}, v_T)]
\end{aligned}$$

that is, $1/T^2$ times the sum of all the variances and all the possible covariance between the terms $v_1, v_2, \ldots, v_T$.

Proof: variance of $\hat{\beta}_1$ with correlated errors V

It is convenient to visualize these variances and covariances in a table:

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \dots & \dots & \dots & \sigma_{1T} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \dots & \dots & \dots & \sigma_{2T} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \dots & \dots & \dots & \sigma_{3T} \\ \dots & & & & & & \\ \sigma_{t1} & \sigma_{t2} & \sigma_{t3} & \dots & \sigma_{tt} & \dots & \sigma_{tT} \\ \dots & & & & & & \\ \sigma_{T1} & \sigma_{T2} & \sigma_{T3} & \dots & \dots & \dots & \sigma_{TT} \end{bmatrix}$$

where σ_{ii} (same index) denotes the variance of v_i and σ_{ij} (different indices) the covariance between v_i and v_j. As σ_{ij} is a covariance, $\sigma_{ij} = \sigma_{ji}$.

In the table, there are:

- T variances (main diagonal), i.e. T terms like σ_{ii}.
- $2(T-1)$ lag-1 covariances, (T-1) above the diagonal and (T-1) below the diagonal (and $\sigma_{ij} = \sigma_{ji}$), i.e. $2(T-1)$ terms like $\sigma_{i(i-1)}$.

Proof: variance of $\hat{\beta}_1$ with correlated errors VI

- $2(T-2)$ lag-2 covariances, (T-2) above and below the second (super- and sub-) (and $\sigma_{ij} = \sigma_{ji}$), i.e. $2(T-2)$ terms like $\sigma_{i(i-2)}$.
- ...
- 2 lag-T covariances, (upper-right and lower-left corners), i.e. 2 terms like σ_{1T}.

Thus,

$$
\begin{aligned}
\mathbb{V}(\bar{v}) = \frac{1}{T^2}[T\,\mathbb{V}(v_t) \\
&+ 2(T-1)\,\mathbb{C}\text{ov}(v_t, v_{t-1}) \\
&+ 2(T-2)\,\mathbb{C}\text{ov}(v_t, v_{t-2}) \\
&+ \dots \\
&+ 2T\,\mathbb{C}\text{ov}(v_t, v_1)] \qquad (8)
\end{aligned}
$$

Proof: variance of $\hat{\beta}_1$ with correlated errors VII

Remember the general definition of correlation for two random variables v_i and v_j:

$$\mathbb{C}\text{ov}(v_i, v_j) \overset{2.26}{=} \mathbb{C}\text{orr}(v_i, v_j)\sqrt{\mathbb{V}(v_i)\,\mathbb{V}(v_j)}$$

but in our setup, (X_i, u_i) are identically distributed over time (joint stationarity), so the variance of $v_i = (X_i - \mu_x)u_i$ is also constant,

$$\mathbb{V}(v_i) = \sigma_v^2, \quad \forall i = 1, \ldots, T$$

so $\mathbb{V}(v_i) = \mathbb{V}(v_j) = \sigma_v^2$, then $\sqrt{\mathbb{V}(v_i)\,\mathbb{V}(v_j)} = \sqrt{(\sigma_v^2)^2} = \sigma_v^2$, therefore

$$\mathbb{C}\text{ov}(v_i, v_j) = \mathbb{C}\text{orr}(v_i, v_j)\sigma_v^2.$$

Proof: variance of $\hat{\beta}_1$ with correlated errors VIII

Then, by multiplying every term in eq.(8) for σ_v^2, we rewrite each covariance as a correlation:

$$\begin{aligned}
\mathbb{V}(\bar{v}) = \frac{1}{T^2}[T\,& \overbrace{\mathbb{C}\text{orr}(v_t, v_t)}^{=1}\sigma_v^2 \\
&+ 2(T-1)\,\mathbb{C}\text{orr}(v_t, v_{t-1})\sigma_v^2 \\
&+ 2(T-2)\,\mathbb{C}\text{orr}(v_t, v_{t-2})\sigma_v^2 \\
&+ \ldots \\
&+ 2\,\mathbb{C}\text{orr}(v_t, v_1)\sigma_v^2],
\end{aligned}$$

Proof: variance of $\hat{\beta}_1$ with correlated errors IX

and by rearranging the above,

$$
\begin{aligned}
\mathbb{V}(\bar{v}) = \frac{1}{T}[\frac{T}{T}\sigma_v^2 \\
&+ 2\frac{(T-1)}{T}\,\mathbb{C}\mathrm{orr}(v_t, v_{t-1})\sigma_v^2 \\
&+ 2\frac{(T-2)}{T}\,\mathbb{C}\mathrm{orr}(v_t, v_{t-2})\sigma_v^2 \\
&+ \dots \\
&+ 2\frac{1}{T}\,\mathbb{C}\mathrm{orr}(v_t, v_1)\sigma_v^2].
\end{aligned}
$$

This is also written as:

$$\begin{aligned}\mathbb{V}(\bar{v}) = \frac{1}{T}[\sigma_v^2 \\ &+ \frac{\sigma_v^2}{T} 2(T-1)\,\mathbb{C}\mathrm{orr}(v_t, v_{t-1}) \\ &+ \frac{\sigma_v^2}{T} 2(T-2)\,\mathbb{C}\mathrm{orr}(v_t, v_{t-2}) \\ &+ \dots \\ &+ \frac{\sigma_v^2}{T} 2\underbrace{(T-T+1)}_{1}\mathbb{C}\mathrm{orr}(v_t, v_1)],\end{aligned}$$

Proof: variance of $\hat{\beta}_1$ with correlated errors XI

and by bringing $1/T$ inside the square bracket,

$$\begin{aligned}\mathbb{V}(\bar{v}) = [&\frac{\sigma_v^2}{T}\\ &+ \frac{\sigma_v^2}{T} 2 \frac{T-1}{T} \mathbb{C}\mathrm{orr}(v_t, v_{t-1})\\ &+ \frac{\sigma_v^2}{T} 2 \frac{T-2}{T} \mathbb{C}\mathrm{orr}(v_t, v_{t-2})\\ &+ \ldots\\ &+ \frac{\sigma_v^2}{T} 2 \frac{T-T+1}{T} \mathbb{C}\mathrm{orr}(v_t, v_1)]\end{aligned}$$

Proof: variance of $\hat{\beta}_1$ with correlated errors XII

and lastly, collecting the common term σ_v^2/T

$$\begin{aligned}\mathbb{V}(\bar{v}) = \frac{\sigma_v^2}{T}[1+ & \\ & + 2\frac{T-1}{T}\,\mathbb{C}\mathrm{orr}(v_t, v_{t-1}) \\ & + 2\frac{T-2}{T}\,\mathbb{C}\mathrm{orr}(v_t, v_{t-2}) \\ & + \dots \\ & + 2\frac{T-T+1}{T}\,\mathbb{C}\mathrm{orr}(v_t, v_1)]\end{aligned}$$

Proof: variance of $\hat{\beta}_1$ with correlated errors XIII

This, in a compact form, is written as

$$\begin{aligned}\mathbb{V}(\bar{v}) &= \frac{\sigma_v^2}{T}\left[1 + 2\sum_{j=1}^{T-1}\frac{T-j}{T}\,\mathbb{C}\mathrm{orr}(v_t, v_{t-j})\right] \\ &= \frac{\sigma_v^2}{T}\left[1 + 2\sum_{j=1}^{T-1}\frac{T-j}{T}\rho_j\right] \\ &= \frac{\sigma_v^2}{T}f_T, \end{aligned} \tag{9}$$

where,

$$f_T = 1 + 2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\rho_j, \qquad \rho_j = \mathbb{C}\mathrm{orr}(v_t, v_{t-j}).$$

Proof: variance of $\hat{\beta}_1$ with correlated errors XIV

Going back to eq.(6),

$$\mathbb{V}\left(\hat{\beta}_1\right) \overset{eq.(6)}{=} \frac{\mathbb{V}(\bar{v})}{\left(\sigma_X^2\right)^2} \overset{eq.(9)}{=} \frac{1}{\left(\sigma_X^2\right)^2}\frac{\sigma_v^2}{T} f_T$$
$$= \underbrace{\left[\frac{1}{T}\frac{\sigma_v^2}{\left(\sigma_X^2\right)^2}\right]}_{eq.(7)} f_T$$

which proves eq.(4).
We conclude that $\mathbb{V}\left(\hat{\beta}_1\right)$ is decomposed into two terms:

i The part in the square brackets, which by eq.(7) is the variance of $\hat{\beta}_1$ as if the errors were serially not correlated.

ii A term f_T that arises when serial correlation is considered (a corrective factor on the first term).

The HAC estimator I

Look at the two terms on the right-hand side of eq.(4), and indicate with:

- $\widehat{\sigma}^2_{\widehat{\beta}_1}$ the estimator of the term $\frac{1}{T}\sigma^2_v/(\sigma^2_X)^2$ (see eq. 5.4)

  $\rightarrow \widehat{\sigma}^2_{\widehat{\beta}_1}$ is the estimator of the variance of $\hat{\beta}_1$ in the absence of serial correlation.
- $\widehat{f}_T$ the estimator of f_T, defined in eq.(5).

The HAC estimator of the variance of $\hat{\beta}_1$, $\mathbb{V}\left(\hat{\beta}_1\right)$ is:

$$\widetilde{\sigma}^2_{\widehat{\beta}_1} = \widehat{\sigma}^2_{\widehat{\beta}_1}\widehat{f}_t \tag{10}$$

This is also called the **Newey-West variance estimator** (in the general case, they showed that along with a rule like eq. (12) the estimator is consistent, when X_t and u_t have more than eight moments, so the earlier assumption).

The HAC estimator II

The task of constructing a consistent estimator $\widehat{f_t}$ is challenging:

- If one replaces the autocorrelations ρ_j with the sample autocorrelations $\hat{\rho}_j$, the estimator $1 + 2\sum_{j+1}^{T-1}\left(\frac{T-j}{T}\right)\hat{\rho}_j$, would intuitively contain so many estimated autocorrelations that it is inconsistent. Because each of the estimated autocorrelations contains an estimation error, by estimating so many autocorrelations the estimation error in this estimator of f_T remains large even in large samples.
- Conversely, using just a few autocorrelations eliminates the problem of estimating several parameters. However, this ignores the autocorrelations at higher lags, leading again to an inconsistent estimator.
- In practice, one balances between the two cases by choosing the number of autocorrelations to include in a way that depends on the sample size T.

The HAC estimator III

Specifically, $\widehat{f}_T$ is given by:

$$\widehat{f}_T = 1 + 2\sum_{j=1}^{m-1}\left(\frac{m-j}{m}\right)\tilde{\rho}_j \tag{11}$$

where,

$$\tilde{\rho}_j = \frac{\sum_{t=j+1}^{T} \hat{v}_t \hat{v}_{t-j}}{\sum_{t=1}^{T} \hat{v}_t^2}, \text{ with } \hat{v} = (X_t - \overline{X})\hat{u}_t.$$

and m is called the **truncation parameter** of the HAC estimator (as it shortens the sum to include $m-1$ terms in place of the $T-1$ in eq. (5)).

The HAC estimator IV

For $\widehat{f}_T$ to be consistent, m must be chosen so that it is large in large samples, although still much less than T, in practice:

$$m = 0.75\,T^{\frac{1}{3}}, \tag{12}$$

rounded to the closest integer.

The HAC estimator V

Notes:

- Newey–West variance estimator is not the only HAC estimator, e.g. weights in eq.(11) can differ and thus alternatives to eq.(12) exist.
- The above can be generalized to regressions with multiple lags and, more generally, to the multiple regression model with serially correlated errors

Estimation of Dynamic Causal Effects with Strictly Exogenous Regressors

Estimation of Dynamic Causal Effects with Strictly Exogenous Regressors

When X_t is strictly exogenous, two alternative estimators of dynamic causal effects are available.

1. The first such estimator involves estimating an ADL model instead of a distributed lag model and calculating the dynamic multipliers from the estimated ADL coefficients.

   $\rightarrow$ this method estimates fewer coefficients than OLS estimation of the distributed lag model, potentially reducing estimation error.

2. The second method is to estimate the coefficients of the distributed lag model, using generalized least squares (GLS) instead of OLS.

   $\rightarrow$ the GLS estimator has a smaller variance.

This first alternative is now discussed in the context of a distributed lag model with a single lag and AR(1) errors.

The Distributed Lag Model with AR(1) Errors I

Suppose that the causal effect on Y of a change in$X$ lasts for only two periods; that is, it has an initial impact effect β_1 and an effect in the next period of β_2 but no effect thereafter.

$$Y_t = \beta_0 + \beta_1 X + \beta_2 X_{t-1} + u_t \tag{13}$$

where u_t is, in general, serially correlated.

As a consequence, inference based on OLS standard errors is misleading.

Two approaches:

1. A solution is to use HAC standard errors
2. If X_t is strictly exogenous, one can adopt an autoregressive model for the serial correlation in u_t and then using this AR model to derive estimators that can be more efficient than OLS

The Distributed Lag Model with AR(1) Errors II

Suppose that u_t follows the AR(1) model,

$$u_t = \phi_1 u_{t-1} + \tilde{u}_t \tag{14}$$

where (i) ϕ_1 is the autoregressive parameter, (ii) $\tilde{u}_t$ are serially uncorrelated, ad (iii) no intercept is needed because $\mathbb{E}(u_t) = 0$.

The model DL eq.(13) with AR(1) errors as of eq.(14) can be written in two ways

- as an ADL model with uncorrelated errors **[PROOF]**
- as a DL model (in some other variables) with uncorrelated errors **[PROOF]**

Proof: The DL model with AR(1) errors has an ADL representation I

To show: the DL model in eq.(13) with AR(1) errors as of eq.(14) can be written as an ADL model.

$\rightarrow$ The goal is thus to rewrite eq.(13) as an ADL model with (a new) serially uncorrelated error (where the dynamics of the autocorrelated error is that of eq.(14)).

Proof: The DL model with AR(1) errors has an ADL representation II

Start with eq.(13) and subtract on both sides $-\phi_1 Y_{t-1}$:

$$\begin{aligned}
Y_t - \phi_1 Y_{t-1} &= \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t - \phi_1 Y_{t-1} \\
&= \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t \\
&\quad - \phi_1(\beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_{t-1}) \\
&= \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} - \phi_1(\beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2}) \\
&\quad + u_t - \phi_1 u_{t-1} \\
&= \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} - \phi_1\beta_0 - \phi_1\beta_1 X_{t-1} - \phi_1\beta_2 X_{t-2} + \tilde{u}_t
\end{aligned} \tag{15}$$

since by eq.(14), $u_t - \phi_1 u_{t-1} = \tilde{u}_t$.

Proof: The DL model with AR(1) errors has an ADL representation III

Rearranging eq. (15),

$$Y_t = \underbrace{\beta_0(1-\phi_1)}_{\alpha_0} + \phi_1 Y_{t-1} + \underbrace{\beta_1}_{\delta_0} X_t + \underbrace{(\beta_2 - \phi_1\beta_1)}_{\delta_1} X_{t-1} - \underbrace{\phi_1\beta_2}_{-\delta_2} X_{t-2} + \tilde{u}_t$$

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \tilde{u}_t \tag{16}$$

we obtain the form of an ADL(1,2) model:

$\rightarrow$ Eq.(16) is the **ADL representation** of eq.(13) with autoregressive errors as of eq.(14).

Proof: quasi-difference representation I

To show: that the DL model in eq.(13) with AR(1) errors as of eq.(14) can be written as another DL model, in some new variables, that has uncorrelated errors.

Going back to eq.(15), reorganize the terms differently:

$$
\begin{aligned}
Y_t - \phi_1 Y_{t-1} &= \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} - \phi_1\beta_0 - \phi_1\beta_1 X_{t-1} - \phi_1\beta_2 X_{t-2} \\
&\quad + \tilde{u}_t \\
\underbrace{Y_t - \phi_1 Y_{t-1}}_{\tilde{Y}} &= \underbrace{\beta_0(1-\phi_1)}_{\alpha_0} + \beta_1 \underbrace{(X_t - \phi_1 X_{t-1})}_{\tilde{X}_t} + \beta_2 \underbrace{(X_{t-1} - \phi_1 X_{t-2})}_{\tilde{X}_{t-1}} + \tilde{u}_t \\
\tilde{Y} &= \alpha_0 + \beta_1 \tilde{X}_t + \beta_2 \tilde{X}_{t-1} + \tilde{u}_t \qquad (17)
\end{aligned}
$$

Eq.(17) is indeed ad DL model, into some new variables, with serially uncorrelated errors.

We refer to eq.(17) as the **quasi-difference representation** of the distributed lag model with autoregressive errors given in eq. (13) and (14).

$\tilde{Y} = Y_t - \phi_1 Y_{t-1}$ is the **quasi-difference** of Y_t because is not the first difference, but the difference between Y_t and $\phi_1 Y_{t-1}$ (same for $\tilde{X}$). Note that the ADL representation in eq.(16) and the quasi-difference representation in eq. (17) are *equivalent*: the error $\tilde{u}_t$ is not serially correlated. The two, however, lead to two different estimation strategies.

Before discussing those strategies, we turn to the assumptions under which they yield consistent estimators of the dynamic multipliers, β_1 and β_2 (of the original problem in eq.(13), with eq.(14)) **[PROOF]**.

Proof: the conditional mean assumption in the ADL and quasi-difference models I

To show: strict exogeneity is needed for estimating the quasi-difference model and the ADL model. **How:** Show that the exogeneity for the quasi-differenced model implies strict exogeneity for the original DL model in X_t, u_t.

Since eq.(16) and eq.(17) are equivalent, the conditions for their estimation are the same, we consider eq.(17).

The quasi-difference model in eq.(17) is a DL model in the quasi-differenced variables with a serially uncorrelated error.
$\rightarrow$ the OLS conditions for estimation are the same as those for the DL model, expressed in terms of $\tilde{X}_t$ and $\tilde{u}_t$.

Proof: the conditional mean assumption in the ADL and quasi-difference models II

In particular, exogeneity is critical:

$$\mathbb{E}\Big(\tilde{u}_t|\tilde{X}_t, \tilde{X}_{t-1}, \ldots\Big) = 0, \tag{18}$$

where letting the conditional expectation depend on distant lags of $\tilde{X}_t$ ensures that no additional lags of $\tilde{X}_t$, other than those appearing in eq.(17), enter the population regression function.

Because $\tilde{X}_t = X_t - \phi X_{t-1}$, conditioning of $\tilde{X}_t$ and all of its lags is equivalent to conditioning on X_t all of its lags. Thus the conditional expectation condition in eq.(18) is equivalent to the condition

$$\mathbb{E}(\tilde{u}_t|X_t, X_{t-1}, \ldots) = 0.$$

Proof: the conditional mean assumption in the ADL and quasi-difference models III

Furthermore, because $\tilde{u}_t = u_t - \phi_1 u_{t-1}$,

$$
\begin{aligned}
0 &= \mathbb{E}(\tilde{u}_t|\tilde{X}_t, \tilde{X}_{t-1}, \dots) \\
&= \mathbb{E}(\tilde{u}_t|X_t, X_{t-1}, \dots) \\
&= \mathbb{E}(u_t - \phi_1 u_{t-1}|X_t, X_{t-1}, \dots) \\
&= \mathbb{E}(u_t|X_t, X_{t-1}, \dots) - \phi\,\mathbb{E}(u_{t-1}|X_t, X_{t-1}, \dots)
\end{aligned}
\tag{19}
$$

For the equality in eq.(19) to hold for *general* (any) values of ϕ_1, it must be the case that both

$$
\mathbb{E}(u_t|X_t, X_{t-1}, \dots) = 0, \qquad \text{and} \qquad \mathbb{E}(u_{t-1}|X_t, X_{t-1}, \dots) = 0.
$$

Proof: the conditional mean assumption in the ADL and quasi-difference models IV

By shifting the time subscripts forward one time period, the condition that $\mathbb{E}(u_{t-1}|X_t, X_{t-1}, \dots) = 0$ can be rewritten as:

$$\mathbb{E}(u_t|X_{t+1}, X_t, X_{t-1}, \dots) = 0 \tag{20}$$

which (by the law of iterated expectations) implies that also $\mathbb{E}(u_t|X_t, X_{t-1}, \dots) = 0$

$\rightarrow$ having the 0 conditional mean assumption in eq.(18) hold for *general* values of ϕ_1 is equivalent to having the condition in eq.(20) hold.
The condition in eq.(20) is implied by X_t being strictly exogenous, but it is not implied by X_t being (past and present) exogenous.

Proof: the conditional mean assumption in the ADL and quasi-difference models V

Thus the least squares assumptions for the estimation of the distributed lag model in eq.(17) hold if X_t is *strictly* exogenous (it is not enough that X_t is past and present exogenous).

Because the ADL representation (eq.(16)) is equivalent to the quasi-differenced representation (eq.(17)), the conditional mean assumption needed to estimate the coefficients of the quasi-differenced representation (strict exogeneity, eq.(20)) is also the conditional mean assumption for consistent estimation of the coefficients of the ADL representation.

OLS Estimation of the ADL Model I

We now turn to the two estimation strategies suggested by these two representations:

1. Estimation of the ADL coefficients (OLS)
2. Estimation of the coefficients of the quasi-difference model (GLS)

The first strategy is to use OLS to estimate the coefficients in the ADL model in eq.(16).

- The error is serially uncorrelated thus no need for HAC standard errors.
- The estimated ADL coefficients are not themselves estimates of the dynamic multipliers, but the dynamic multipliers can be computed from the ADL coefficients **[PROOF]**.

Proof: dynamic multipliers estimation I

To show: How to estimate the dynamic multipliers of the DL model using the ADL representation.

A general way to compute the dynamic multipliers is to express the estimated regression function as a function of current and past values of X_t, i.e., to eliminate Y_t from the estimated regression function. To do so, repeatedly substitute expressions for lagged values of Y_t into the estimated regression function.

Consider the estimated regression function of eq.(16):

$$\hat{Y}_t = \hat{\phi}_1 Y_{t-1} + \hat{\delta}_0 X_t + \hat{\delta}_1 X_{t-1} + \hat{\delta}_2 X_{t-2} \tag{21}$$

where the estimated intercept has been omitted because it does not enter any expression for the dynamic multipliers.

Proof: dynamic multipliers estimation II

Lagging both sides of eq.(21) yields,

$$\hat{Y}_{t-1} = \hat{\phi}_1 Y_{t-2} + \hat{\delta}_0 X_{t-1} + \hat{\delta}_1 X_{t-2} + \hat{\delta}_2 X_{t-3} \tag{22}$$

that, used in eq.(21) gives:

$$\begin{aligned}
\hat{Y}_t &= \hat{\phi}_1 \Big(\hat{\phi}_1 Y_{t-2} + \hat{\delta}_0 X_{t-1} + \hat{\delta}_1 X_{t-2} + \hat{\delta}_2 X_{t-3} \Big) \\
&\quad + \hat{\delta}_0 X_t + \hat{\delta}_1 X_{t-1} + \hat{\delta}_2 X_{t-2} \\
&= \hat{\delta}_0 X_t + \Big(\hat{\delta}_1 + \hat{\phi}_1 \hat{\delta}_0 \Big) X_{t-1} + \Big(\hat{\delta}_2 + \hat{\phi}_1 \hat{\delta}_1 \Big) X_{t-2} + \hat{\phi}_1 \hat{\delta}_2 X_{t-3} \\
&\quad + \hat{\phi}_1^2 Y_{t-2}
\end{aligned} \tag{23}$$

Now repeat the process repeatedly substituting expressions for Y_{t-2}, Y_{t-3}, ... (e.g., for Y_{t-2} lag eq.(22) and use it in eq.(23)).

Proof: dynamic multipliers estimation III

This gives:

$$\begin{aligned}
\hat{Y}_t = {} & \hat{\delta}_0 X_t + \left(\hat{\delta}_1 + \hat{\phi}_1 \hat{\delta}_0\right) X_{t-1} \\
& + \left(\hat{\delta}_2 + \hat{\phi}_1 \hat{\delta}_1 + \hat{\phi}_1^2 \hat{\delta}_0\right) X_{t-2} \\
& + \hat{\phi}_1 \left(\hat{\delta}_2 + \hat{\phi}_1 \hat{\delta}_1 + \hat{\phi}_1^2 \hat{\delta}_0\right) X_{t-3} \\
& + \hat{\phi}_1^2 \left(\hat{\delta}_2 + \hat{\phi}_1 \hat{\delta}_1 + \hat{\phi}_1^2 \hat{\delta}_0\right) X_{t-4} + \dots
\end{aligned} \tag{24}$$

The coefficients in eq.(24) are the estimators of the dynamic multipliers, computed from the OLS estimators of the coefficients in the ADL model in eq.(16).

Proof: dynamic multipliers estimation IV

Look at the restrictions in eq.(16):

$$\alpha_0 = \beta_0(1-\phi_1), \quad \delta_0 = \beta_1, \quad \delta_1 = \beta_2 - \phi_1\beta_1, \quad \delta_2 = -\phi_1\beta_2 \tag{25}$$

If these were to hold exactly for the estimated coefficients, then the dynamic multipliers beyond the second one (that is, the coefficients on $X_{t-2}, X_{t-3,\ldots,}$) would all be 0.

In fact, the coefficients of $X_{t-2}, X_{t-3,\ldots}$ always involve the term $\hat{\delta}_2 + \hat{\phi}_1\hat{\delta}_1 + \hat{\phi}_1^2\hat{\delta}_0$, that, from eq.(25), is (should be) zero:

$$\begin{aligned}
\delta_2 + \phi_1\delta_1 + \phi_1^2\delta_0 &= -\phi_1\beta_2 + \phi_1(\beta_2 - \phi_1\beta_1) + \phi_1^2\beta_1 \\
&= -\phi_1\beta_2 + \phi_1\beta_2 - \phi_1^2\beta_1 + \phi_1^2\beta_1 \\
&= 0
\end{aligned}$$

However, under this estimation strategy, these restrictions will not hold exactly, so the estimated multipliers beyond the second in eq.(24) will generally be nonzero.

ADL estimation, summary

In conclusion, from the proof, to estimate the dynamic multipliers of the DL model in eq.(13) with AR(1) errors as for eq.(14):

$$Y_t = \beta_0 + \beta_1 X_{t-1} + u_t, \qquad \text{where} \quad u_t = \phi_1 u_{t-1} + \tilde{u}_t,$$

under the ADL representation in eq.(16):

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \tilde{u}_t$$

- Estimate the above with OLS, and obtain $\hat{\phi}_1, \hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_2$
- From eq.(24),
    - The first dynamic multiplier is $\hat{\delta}_0$
    - The second dynamic multiplier is $\hat{\delta}_1 + \hat{\phi}_1 \hat{\delta}_0$

GLS Estimation I

How about the estimation of the quasi-differenced representation in eq.(17)?

If ϕ_1 is known, then the quasi-differenced variables $\tilde{X}_t$ and $\tilde{Y}_t$ can be computed directly, and if X_t is *strictly* exogenous the coefficients in the representation quasi-difference representation

$$\tilde{Y} = \alpha_0 + \beta_1 \tilde{X}_t + \beta_2 \tilde{X}_{t-1} + \tilde{u}_t$$

can be computed directly with OLS.

In reality, this is an infeasible approach because ϕ_1 is unknown, so $\tilde{X}_t$ and $\tilde{Y}_t$ cannot be computed and thus the OLS estimator cannot be implemented.

This approach is called the **infeasible generalized least squares (GLS) estimator**.

GLS Estimation II

The **feasible GLS estimator** modifies the infeasible GLS estimator by using a preliminary estimator of ϕ_1, $\hat{\phi}_1$, to compute the estimated quasi-differences.

This is how it works:

- Estimate the DL regression in eq.(13) by OLS

$$Y_t = \beta_0 + \beta_1 X_{t-1} + u_t,$$

  from which one obtains $\hat{\beta}_0$ and $\hat{\beta}_1$, and thus the residuals $\hat{u}_t$.
- Estimate ϕ by replacing u_t in eq.(14) with the residuals $\hat{u}_t$. That is, run the OLS regression

$$\hat{u}_t = \phi \hat{u}_{t-1} + \tilde{u}_t$$

  and obtain $\hat{\phi}$.

GLS Estimation III

- With the estimate $\hat{\phi}_1$ in hand, compute the estimated quasi-differences of Y_t, X_t and X_{t-1}:

$$\hat{\tilde{Y}}_t = Y_t - \hat{\phi}_1 Y_{t-1},$$

$$\hat{\tilde{X}}_t = X_t - \hat{\phi}_1 X_{t-1}, \qquad \hat{\tilde{X}}_{t-1} = X_{t-1} - \hat{\phi}_1 X_{t-2},$$

- Obtain the feasible GLS estimator of eq.(17) by regressing $\hat{\tilde{Y}}_t$ on $\hat{\tilde{X}}_t$ and $\hat{\tilde{X}}_{t-1}$. That is, estimate with OLS

$$\hat{\tilde{Y}}_t = \alpha_0 + \beta_1 \hat{\tilde{X}}_t + \beta_2 \hat{\tilde{X}}_{t-1} + \tilde{u}_t \tag{26}$$

The OLS estimates of eq.(26) estimate the coefficients in the quasi-differenced representation eq.(17), and so the coefficients of the initial the DL model in eq.(13) with AR(1) errors (eq.(14)).

GLS Estimation IV

Efficiency of the GLS estimator:

- If $\hat{u}$ is homoskedastic, if ϕ_1 is known and if X_t is strictly exogenous: the OLS estimator of $\alpha_0, \beta_1, \beta_2$ in eq.(17) is BLUE.

  Specifically is the best estimator among all the linear conditionally unbiased estimators based on $\tilde{X}_t$ and $\tilde{Y}_t$ for $t = 2, \ldots, T$, where the first observation $(t = 1)$ is lost because of quasi-differencing.
- In the feasible GLS estimator ϕ_1 is estimated. Because the estimator of ϕ_1 is consistent and its variance is inversely proportional to T, the feasible and infeasible GLS estimators have the same variances in large samples, and the loss of information from the first observation $(t = 1)$ is negligible when T is large.

GLS Estimation V

- Therefore, if X is strictly exogenous, then also the feasible GLS estimator is BLUE in large samples
- Furthermore, if X is strictly exogenous, then GLS is more efficient than the OLS estimator of the distributed lag coefficients in eq.(13).
- (if X is only past and present exogenous the GLS estimator is not consistent)