Distributed lag models *

Jean-Marie Dufour †
McGill University

April 14, 2010, 1:30pm

* This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

† William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

Contents

1. Models with lagged explanatory variables

It is quite common to explain a variable using current and past values of another variable, as in the following equations:

$$I_t = \alpha + \beta_0 V_t + \beta_1 V_{t-1} + \beta_2 V_{t-2} + \cdots + \beta_k V_{t-k} + u_t \tag{1.1}$$

where I_t represents investment (t time t) and V_t represents sales at time t;

$$\hat{p}_t = \alpha + \beta_0 \hat{M}_t + \beta_1 \hat{M}_{t-1} + \cdots + \beta_k \hat{M}_{t-k} + u_t \tag{1.2}$$

where $\hat{p}_t$ is the rate of inflation and $\hat{M}_t$ is the growth rate of a monetary aggregate. If we suppose that

$$u_t \overset{ind}{\sim} N\left[0, \sigma^2\right], \; t = 1, \ldots, T, \tag{1.3}$$

these equations can be estimated by ordinary least squares.

However, if k is large, one typically has a large number of strongly collinear explanatory variables. Restrictions on the coefficients $\beta_i, i = 0, 1, \ldots, k$, are typically needed to improve estimation precision. Several approaches have been suggested.

Consider a general equation of the form:

$$y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \ldots + \beta_k X_{t-k} + u_t = \alpha + \sum_{i=0}^{k} \beta_1 X_{t-i} + u_t \,.$$

(1) Arithmetic progression (I. Fisher, 1937):

$$\begin{aligned} \beta_i &= (k+1-i)\,\beta\,, \quad 0 \leq i \leq k \\ &= 0\,, \qquad\qquad\quad\; i > k \end{aligned} \tag{1.4}$$

$$\begin{aligned} y_t &= \alpha + \sum_{i=0}^{k} \beta_i X_{t-i} + u_t \\ &= \alpha + \left[\sum_{i=0}^{k} (k+1-i)\, X_{t-i}\right] \beta + u_t \quad , \quad t = 1, \ldots, T. \end{aligned} \tag{1.5}$$

(2) Inverted V progression (De Leeuw, 1962):

$$\begin{aligned} \beta_i &= i\beta \quad , \quad 0 \leq i \leq k/2 \\ &= (k-i)\,\beta \quad , \quad k/2 \leq i \leq k \end{aligned} \tag{1.6}$$

$$y_t = \alpha + \beta \left[\sum_{i=0}^{k/2} X_{t-i} + \sum_{(k/2)+1}^{k} (k-i)\, X_{t-i} \right] + u_t \,. \tag{1.7}$$

(3) Polynomial progression (Almon, 1965):

$$\beta_i = a_0 + a_1 i + a_2 i^2 + \cdots + a_r i^r \,, \qquad 0 \le r < k \,. \tag{1.8}$$

For example, if $r = 2$ and $k = 3$,we get:

$$y_t = \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t \tag{1.9}$$

where

$$\begin{aligned} \beta_t &= a_0 + a_1 t + a_2 t^2 \,, \\ \beta_0 &= a_0 \,, \\ \beta_1 &= a_0 + a_1 + a_2 \,, \\ \beta_2 &= a_0 + 2a_1 + 4a_2 \,, \\ \beta_3 &= a_0 + 3a_1 + 9a_2 \,, \end{aligned}$$

hence

$$y_t = a_0 \left(\sum_{i=0}^{3} X_{t-i} \right) + a_1 \left[\sum_{i=0}^{3} i X_{t-i} \right] + a_2 \left[\sum_{i=0}^{3} i^2 X_{t-i} \right] . \tag{1.10}$$

The number of parameters goes from 4 to 3.

$$\begin{gathered} \beta_{-1} = \beta_{k+1} = 0 \,, \\ \beta_{k+1} = 0 \,, \\ \beta_{-1} = 0 \,. \end{gathered}$$

For example,

$$\beta_{-1} = 0 = a_0 - a_1 + a_2 \Longrightarrow a_2 = a_1 - a_0 \,. \tag{1.11}$$

The Almon method produces very regular patterns for β_i, but we should not forget this is due to the imposed restrictions.

One can test whether a lagged variable is present (given k). If

$$y_t = \alpha + \sum_{i=0}^{k} \beta_i X_{t-i} + \sum_{j=0}^{\ell} \gamma_j Z_{t-j} + u_t \,,$$

one can assume two different polynomials for β_i and γ_j.

Another possibility consists in assuming that

$$\beta_i = \alpha_i + \gamma_i W_{t-i}$$

hence

$$y_t = \alpha + \sum_{i=0}^{k} \alpha_i X_{t-i} + \sum_{i=0}^{k} \gamma_i W_{t-i} X_{t-i} + u_t$$

with two different on α_i and γ_j.

(4) Geometric progression (Koyck, 1954)

$$y_t = \alpha + \beta \sum_{i=0}^{\infty} W_i X_{t-i} + u_t$$

where

$$W_i = (1-\lambda)\,\lambda^i\ ,\ 0 < \lambda < 1,\ i = 0, 1, \ldots \tag{1.12}$$

This leads to the following model:

$$\begin{aligned} y_t &= \alpha + \beta\,(1-\lambda) \sum_{i=0}^{\infty} \lambda^i X_{t-i} + u_t \\ &= \alpha + \beta\,(1-\lambda) \sum_{i=0}^{\infty} \lambda^i L^i X_t + u_t \end{aligned} \tag{1.13}$$

where the lag operator L is defined as follows:

$$\begin{aligned} LX_t &= X_{t-1}\,, \\ L^2 X_t &= L\,(LX_t) = L\,(X_{t-1}) = X_{t-2}\,, \\ L^n X_t &= X_{t-n}\,. \end{aligned}$$

The model then becomes:

$$\begin{aligned} y_t &= \alpha + \beta\,(1-\lambda) \sum_{i=0}^{\infty} (\lambda L)^i\, X_t + u_t \\ &= \alpha + \beta\,(1-\lambda) \left[\sum_{i=0}^{\infty} (\lambda L)^i\right] X_t + u_t \\ &= \alpha + \beta\,(1-\lambda)\,\frac{1}{1-\lambda L} X_t + u_t \end{aligned} \tag{1.14}$$

hence

$$\begin{aligned}(1-\lambda L)\, y_t &= (1-\lambda L)\, \alpha + \beta\,(1-\lambda)\, X_t + (1-\lambda L)\, u_t \\ y_t - \lambda y_{t-1} &= (1-\lambda)\, \alpha + \beta\,(1-\lambda)\, X_t + u_t - \lambda u_{t-1}\end{aligned}$$

or

$$y_t = (1-\lambda)\, \alpha + \lambda y_{t-1} + \beta\,(1-\lambda)\, X_t \quad + \quad \underset{\substack{\uparrow \\ \text{Moving average}}}{(u_t - \lambda u_{t-1})} \quad t = 2, \ldots, T\,.$$

This provides a considerable reduction of the number of parameters. However there is a lagged dependent variable on the right-hand side with autocorrelated errors:
Y_{t-1} is not independent of $u_t - \lambda u_{t-1}$.

We may also wish not to impose the geometric scheme on the first β_i coefficients. For example, if we do this for the first two coefficients, we get:

$$\begin{aligned}y_t &= \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 \sum_{i=2}^{\infty} \lambda^i X_{t-i} \; + \; u_t \\ &= \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 \frac{\lambda^2}{1-\lambda L} X_{t-2} \; + \; u_t \end{aligned} \tag{1.15}$$

or

$$y_t = \alpha\,(1-\lambda) + \lambda y_{t-1} + \beta_0 X_t + (\beta_1 - \lambda\beta_0)\, X_{t-1} + (\beta_2 - \lambda\beta_1)\, X_{t-2} + u_t. \tag{1.16}$$

Another possibility consists in considering geometric progressions on two explanatory variables:

$$\begin{aligned}y_t &= \alpha + \frac{\beta\,(1-\lambda)}{1-\lambda L} X_t + \frac{\gamma\,(1-\lambda)}{1-\lambda L} Z_t + u_t\,, \\ y_t &= \alpha\,(1-\lambda) + \beta\,(1-\lambda)\, X_t + \gamma\,(1-\lambda)\, Z_t + (u_t - \lambda u_{t-1})\ .\end{aligned}$$

Other proposed schemes include: Pascal, rational (Jorgenson), Gamma.

2. Models with lagged dependent variables: economic examples

2.1. Partial adjustment model (Nerlove, 1958)

$$Y_t^* = \alpha + \beta X_t \qquad \text{Equilibrium equation} \tag{2.1}$$

$$Y_t - Y_{t-1} = \gamma \left(Y_t^* - Y_{t-1}\right) + u_t\,, \qquad \text{Adjustment equation} \tag{2.2}$$

$$0 < \gamma \leq 1$$

$$u_t \overset{ind}{\sim} N\left[0, \sigma_u^2\right]$$

$$\begin{aligned} Y_t &= Y_{t-1} + \gamma\left[\alpha + \beta X_t - Y_{t-1}\right] + u_t \\ Y_t &= \alpha\gamma + (1-\gamma) Y_{t-1} + \gamma\beta X_t + u_t \end{aligned} \tag{2.3}$$

Equation (2.2) can be obtained as the solution a cost minimization problem:

$$C_t = \underset{\text{Disequilibrium cost}}{\underset{\uparrow}{a\left(Y_t - Y_t^*\right)}} + \underset{\text{Adjustment cost}}{\underset{\uparrow}{b\left(Y_t - Y_{t-1}\right)^2}} \quad , \quad a \geq 0,\ b \geq 0 \quad , \quad a + b \neq 0\,. \tag{2.4}$$

Differentiating (2.4) with respect to Y_t, we get:

$$\begin{aligned} \frac{\partial C_t}{\partial Y_t} = \ & 2a\left(Y_t - Y_t^*\right) + 2b\left(Y_t - Y_{t-1}\right) = 0 \\ \Longrightarrow \ & (a+b)\, Y_t - aY_t^* - bY_{t-1} = 0 \\ & (a+b)\, Y_t - (a+b)\, Y_{t-1} + aY_{t-1} - aY_t^* = 0 \\ & Y_t - Y_{t-1} = \tfrac{a}{a+b}\left(Y_t^* - Y_{t-1}\right) \\ & \qquad\qquad = \gamma\left(Y_t^* - Y_{t-1}\right) \end{aligned}$$

$$0 \leq \frac{a}{a+b} \leq 1.$$

2.2. Adaptive expectations

$$\begin{aligned} Y_t &= \alpha + \beta X_t^* + u_t,\ t = 1, \ldots, T \\ & u_t \overset{ind}{\sim} N\left[0, \sigma_u^2\right] \end{aligned} \tag{2.5}$$

$$
\begin{aligned}
X_t^* - X_{t-1}^* &= \delta\left(X_{t-1} - X_{t-1}^*\right) && \text{Expectation equation} \qquad (2.6)\\
0 &< \delta < 1 \\
X_t^* &= \text{expected level of } X_t
\end{aligned}
$$

$$
\begin{aligned}
X_t^* - (1-\delta)\,X_{t-1}^* &= \delta X_{t-1} && B = L\\
\left[1-(1-\delta)\,B\right]X_t^* &= \delta X_{t-1}\\
X_t^* &= \frac{\delta}{1-(1-\delta)\,B}X_{t-1}\\
Y_t &= \alpha + \beta\frac{\delta}{1-(1-\delta)\,B}X_{t-1} + u_t\\
\left[1-(1-\delta)\,B\right]Y_t &= \left[1-(1-\delta)\,B\right]\alpha + \beta\delta X_{t-1} + \left[1-(1-\delta)\,B\right]u_t\\
Y_t - (1-\delta)\,Y_{t-1} &= \alpha\delta + \beta\delta X_{t-1} + u_t - (1-\delta)\,u_{t-1}
\end{aligned}
$$

$$
\begin{aligned}
Y_t &= \alpha\delta + \lambda Y_{t-1} + \beta\delta X_{t-1} + (u_t - \lambda u_{t-1}) \qquad (2.7)\\
\lambda &= 1-\delta.
\end{aligned}
$$

3. Estimation of models with lagged dependent variables

Consider an equation of the type:

$$
Y_t = \beta_0 + \beta_1 Y_{t-1} + \underset{\substack{\downarrow \\ \text{Non-stochastic}}}{\beta_2 X_t} + v_t\,,\ t = 0, 1, \ldots, T,
$$

with three alternative hypotheses on the errors:

$$
\begin{aligned}
&\text{(I)} \quad v_t \overset{ind}{\sim} N\left[0, \sigma_v^2\right]\\
&\text{(II)} \quad v_t = u_t - \lambda u_{t-1}\\
&\qquad (a) \quad u_t \overset{ind}{\sim} N\left[0, \sigma_u^2\right] \qquad (3.1)\\
&\qquad (b) \quad u_t = \rho u_{t-1} + \varepsilon_t\\
&\text{(III)} \quad v_t = \rho v_{t-1} + \varepsilon_t
\end{aligned}
$$

3.1. Hypothesis I

$$L = P(Y_1, \ldots, Y_t \mid Y_0, X) = \frac{1}{(2\Pi\sigma_v^2)^{T/2}} \exp\left\{-\frac{1}{2\sigma_v^2}\sum_{t=1}^{T}(Y_t - \beta_0 - \beta_1 Y_{t-1} - \beta_2 X_t)^2\right\}$$

$$Max\, L \implies Min \sum_{t=1}^{T}(Y_t - \beta_0 - \beta_1 Y_{t-1} - \beta_2 X_t)^2$$

$\rightarrow$ MCO are valid $\langle$ Consistent
Asymptotically normal, efficient

(Tests and confidence intervals are only approximate).

The problem gets more complicated when v_t and Y_{t-1} are correlated. Consider:

$$Y_t = \beta Y_{t-1} + v_t$$

$$v_t = \rho v_{t-1} + \varepsilon_t\,, \qquad \varepsilon_t \overset{ind}{\sim} N\left[0, \sigma_\varepsilon^2\right]\,, \quad |\rho| < 1$$

$$\hat{\beta} = \frac{\sum_{t=1}^{T} Y_t Y_{t-1}}{\sum_{t=2}^{T} Y_{t-1}^2}$$

$$Y_t = \beta Y_{t-1} + \rho v_{t-1} + \varepsilon_t$$

$$v_{t-1} = Y_{t-1} - \beta Y_{t-2}$$

$$Y_t = \beta Y_{t-1} + \rho(Y_{t-1} - \beta Y_{t-2}) + \varepsilon_t$$

$$= (\beta + \rho) Y_{t-1} - \beta\rho Y_{t-2} + \varepsilon_t$$

$$\sum_{t=2}^{T} Y_t Y_{t-1} = (\beta + \rho)\sum_{t=2}^{T} Y_{t-1}^2 - \beta\rho\sum_{t=2}^{T} Y_{t-1}Y_{t-2} + \sum_{t=1}^{T}\varepsilon_t Y_{t-1}$$

$$\hat{\beta} = (\beta + \rho) - \beta\rho\,\frac{\sum_{t=2}^{T} Y_{t-1}Y_{t-2}}{\sum_{t=2}^{T} Y_{t-1}^2} + \frac{\sum_{t=1}^{T}\varepsilon_t Y_{t-1}}{\sum_{t=2}^{T} Y_{t-1}^2}$$

$$p\lim \frac{\sum_{t=1}^{T}\varepsilon_t Y_{t-1}}{\sum_{t=2}^{T} Y_{t-1}^2} = 0$$

$$p\lim \hat{\beta} = (\beta + \rho) - \beta\rho\, p\lim \hat{\beta}$$

$$p\lim \hat{\beta} = \frac{\beta + \rho}{1 + \beta\rho}$$

$$p\lim \hat{\beta} - \beta = \rho\left(1 - \beta^2\right) / (1 + \beta\rho)$$

The bias $p\lim \hat{\beta} - \beta$ can be large when $\beta \simeq 0.5$ or $\rho \simeq 0.5$. Similarly, the Durbin-Watson test is not valid:

$$\begin{aligned} d &= \sum_{t=2}^{T} \left(\hat{v}_t - \hat{v}_{t-1}\right)^2 / \sum_{t=1}^{T} \hat{v}_t^2 \\ &= \frac{1}{\sum_{t=1}^{T} \hat{v}_t^2} \left[\sum_{t=2}^{T} \hat{v}_t^2 + \sum_{t=2}^{T} \hat{v}_{t-1}^2 - 2 \sum_{t=1}^{T} \hat{v}_t \hat{v}_{t-1} \right] \\ p \lim d &= 2 - 2\, p \lim \frac{\sum_{t=2}^{T} \hat{v}_t \hat{v}_{t-1}}{\sum_{t=1}^{T} \hat{v}_t^2} \end{aligned}$$

$$\begin{aligned} r_1 &= \sum_{t=2}^{T} \hat{v}_t \hat{v}_{t-1} / \sum_{t=1}^{T} \hat{v}_t^2 \\ \hat{v}_t &= Y_t - \hat{\beta} Y_{t-1} \qquad (\text{if } Y_t = \beta Y_{t-1} + v_t) \\ \hat{v}_{t-1} &= Y_{t-1} - \hat{\beta} Y_{t-2}. \end{aligned}$$

$$\begin{aligned} p \lim\, r_1 &= \frac{\beta\rho\left(\beta + \rho\right)}{1 + \beta\rho} = \rho - \frac{\rho\left(1 - \beta^2\right)}{1 + \beta\rho} \\ &= \frac{\rho\left(1 + \beta\rho\right) - \rho\left(1 - \beta^2\right)}{1 + \beta\rho} = \frac{\beta\rho^2 - \rho\beta^2}{1 + \beta\rho} \\ p \lim\, d &= 2\left[1 - \frac{\beta\rho\left(\beta + \rho\right)}{1 + \beta\rho}\right] . \end{aligned}$$

If we knew the true errors v_t, we could compute:

$$\begin{aligned} d^* &= \sum_{t=2}^{T} \left(v_t - v_{t-1}\right)^2 / \sum_{t=1}^{T} v_t^2 \\ &= \left[\sum_{t=2}^{T} v_t^2 + \sum_{t=2}^{T} v_{t-1}^2 - 2 \sum_{t=2}^{T} v_t v_{t-1} \right] / \sum_{t=1}^{T} v_t^2 \\ p \lim d^* &= 2\left(1 - \rho\right) \\ p \lim \left(d - d^*\right) &= \frac{2\rho\left(1 - \beta^2\right)}{1 + \beta\rho} = 2 p \lim \left(\hat{\beta} - \beta\right) \\ & \quad d \text{ is biased upward } (NS). \end{aligned}$$

A possible solution consists in using Durbin's test:

$$\begin{aligned} r &= \sum_{t=2}^{T} \hat{v}_t \hat{v}_{t-1} / \sum_{t=1}^{T-1} \hat{v}_t^2 \simeq 1 - \frac{1}{2} d \\ h &= r \sqrt{\frac{n}{1 - n\hat{V}(b_1)}} \end{aligned}$$

where $n = T - 1$ and $\hat{V}(b_1)$ is the estimator of the variance of b_1, and β_1 is the coefficient de Y_{t-1}.Then

$$h \sim N[0,1] \text{ under } H_0.$$

This test is applicable even if $Y_{t-2}, Y_{t-3}, \ldots$ also appear in the equation. The test is not applicable if $n\hat{V}(b_1) \geq 1$.

In view of then latter fact, an alternative test is based on considering a regression of the type:

$$\hat{v}_t = \gamma \hat{v}_{t-1} + \text{ other regressors, } t = 2, \ldots, T.$$

We then test $\gamma = 0$ using a standard t standard, e.g.

$$\hat{v}_t = \gamma \hat{v}_{t-1} + \gamma_0 + \gamma_1 Y_{t-1} + \gamma_2 X_t + e_t, \ t = 2, ..., T \ .$$

3.2. Hypothesis IIa

$$\begin{aligned} &v_t = u_t - \lambda u_{t-1} && \text{Koyck scheme} \\ &u_t \overset{ind}{\sim} N[0, \sigma_n^2] && \text{Adaptive expectations} \end{aligned}$$

$$\begin{aligned} E(v_t) &= 0, \forall t \\ E(v_t^2) &= \sigma_u^2 \left(1 + \lambda^2\right), \forall t \\ E[v_t v_{t+s}] &= -\lambda \sigma_u^2, \ s = \pm 1, \forall t \\ &= 0 \text{ , pour } |s| \geq 2, \forall t \end{aligned}$$

$$V = E[v\, v'] = \sigma_u^2 \begin{bmatrix} 1+\lambda^2 & -\lambda & 0 & \ldots & 0 \\ -\lambda & 1+\lambda^2 & -\lambda & \ldots & 0 \\ \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & \ldots & 1+\lambda^2 \end{bmatrix} .$$

In Koyck models or with adaptive expectations, we have:

$$\begin{aligned} \beta_1 &= \lambda \\ Y_t &= \beta_0 + \lambda Y_{t-1} + \beta_2 X_t + v_t \end{aligned}$$

$$
\begin{aligned}
y_t &= Y_t - \lambda Y_{t-1} = \beta_0 + \beta_2 X_t + v_t \\
\hat{\beta} &= \left(X'V^{-1}X\right)^{-1} X'V^{-1}y \qquad y = \begin{pmatrix} y_1 \\ \vdots \\ y_t \end{pmatrix}.
\end{aligned}
$$

From a more practical viewpoint,

$$
\begin{aligned}
W_t &= Y_t - u_t \\
W_t &= \lambda W_{t-1} + \beta_0 + \beta_2 X_t \\
W_t &= \lambda \left[\lambda W_{t-2} + \beta_0 + \beta_2 X_{t-1}\right] + \beta_0 + \beta_2 X_t \\
&= \lambda^2 W_{t-2} + \beta_0 (1+\lambda) + \beta_2 (X_t + \lambda X_{t-1})
\end{aligned}
$$

$$
\begin{aligned}
Y_t &= \lambda^t W_0 + \beta_0 \left(1 + \lambda + \lambda^2 + \ldots + \lambda^{t-1}\right) \\
&\quad + \beta_2 (X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} \\
&\quad + \ldots + \lambda^{t-1} X_1) + u_t. \\
&\quad W_0 \text{ is a parameter.}
\end{aligned}
$$

We choose a grid of values of λ over the admissible interval $0 \le \lambda < 1$.Then we compute a linear regression and select the value of λ which minimizes the residual sum of squares [Zellner and Geisel (1968)]. This is asymptotically equivalent to applying maximum likelihood.

3.3. Hypothesis IIb

$$
\begin{aligned}
v_t &= u_t - \lambda u_{t-1} \\
u_t &= \rho u_{t-1} + \varepsilon_t \qquad |\rho| < 1, \qquad \varepsilon_t \overset{ind}{\sim} N\left[0, \sigma^2\right] \\
Y_t &= \beta_0 + \lambda Y_{t-1} + \beta_2 X_t + u_t - \lambda u_{t-1} \\
\rho Y_{t-1} &= \rho \beta_0 + \lambda \rho Y_{t-2} + \beta_2 \rho X_{t-1} + \rho u_{t-1} - \lambda \rho u_{t-2} \\
Y_t - \rho Y_{t-1} &= \beta_0 (1-\rho) + \lambda \left[Y_{t-1} - \rho Y_{t-2}\right] + \beta_2 \left[X_t - \rho X_{t-1}\right] + \varepsilon_t - \lambda \varepsilon_{t-1} \\
Y_t(\rho) &= \beta_0 (1-\rho) + \lambda Y_{t-1}(\rho) + \beta_2 X_t(\rho) + \varepsilon_t - \lambda \varepsilon_{t-1}.
\end{aligned}
$$

Proceeding as for IIa, we define:

$$
\begin{aligned}
Y_t(\rho) &= Y_t - \rho Y_{t-1} \\
X_t(\rho) &= X_t - \rho X_{t-1}
\end{aligned}
$$

$$W_t(\rho) = W_t - \rho W_{t-1}$$

$$\begin{aligned} Y_t(\beta) &= \lambda^t W_0(\rho) + \beta_0 (1-\rho)\left[1 + \lambda + \ldots + \lambda^{t-1}\right] \\ &\quad + \beta_2 \left[X_t(\rho) + \lambda X_{t-1}(\rho) + \ldots + \lambda^{t-1} X_1(\rho)\right] + \varepsilon_t. \end{aligned}$$

We select a grid over

$$\begin{aligned} -1 &\leq \rho \leq 1 \\ 0 &\leq \lambda < 1. \end{aligned}$$

3.4. Hypothesis III

$$\begin{aligned} v_t &= \rho v_{t-1} + \varepsilon_t \quad |\rho| < 1 \quad , \quad \varepsilon_t \sim N\left[0, \sigma^2\right] \\ Y_t &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_t + v_t \\ Y_t - \rho Y_{t-1} &= \beta_0 (1-\rho) + \beta_1 (Y_{t-1} - \rho Y_{t-2}) + \beta_2 (X_t - \rho X_{t-1}) + \varepsilon_t. \end{aligned}$$

This can be estimated by applying Cochrane-Orcutt or the Hildreth-Lu algorithm. Without such transformations, least squares estimators would be inconsistent.