A BUSINESS CYCLE STUDY
FOR THE U.S. FROM 1889 TO 1982

Carlos Ivan Simonsen Leal*

* EPGE/FGV. I am very grateful to John Taylor for helpful discussions. I am responsible for all possible errors.
1. INTRODUCTION

The main task of the economists that study the behaviour of business cycles is to explain why and how certain economic variables tend to fluctuate away from their expected trends. They put questions such as why unemployment was preceded by an unexpected fall in demand.

Our objective in this paper is far more modest. We set up, solve and estimate a very simple rational-expectations model for the U.S. economy. We assumed that all the variables involved followed some linear trend which we will not be interested in estimating. From a statistical point of view it made sense to divide the data into three groups: 1) from 1889 to 1914; 2) from 1909 to 1940, and 3) from 1953 to 1982. The World War II years and the seven following years were ignored taking into consideration that it was a period of heavily constrained demand and post-war adjustment.

There is a problem of high collinearity between nominal wages and prices during the Great Depression. This fact will imply that we will not be able to estimate all the coefficients in our model. Estimation is done according to the techniques developed in the appendix. There the reader will find an extended Cramer - Rao inequality, which we employ to study the asymptotic distribution of the maximum likelihood estimators.
2. THE MODEL

Our aim is to set up and estimate a simultaneous equations rational-expectations model. We will work with three variables: product, wage and price level.

The notation is as follows: capital letters indicate nominal values and small letters are real values. All quantities, unless otherwise stated, are expressed in logarithms. $Y$ stands for product, $P$ price level, $W$ for wage, $M$ for money and $t$ for time (not in logs). The use of a $D$ means we are taking the first difference, for example $DP_t = P_{t+1} - P_t$. $E_t$ indicates expectation conditional on the information available at time $t$ and, as a matter of convention, we assume that all variables with subscripts less than or equal to $t$ are contained in the information set at time $t$.

Inspired in Barro [1], we assume that the supply of product obeys

\[ y_t = \alpha_0 + \alpha_1 t + \alpha_2 (P_t - E_t P_{t+1}) + \eta_t \]

\[ \eta_t = \sigma_t + \theta_1 \sigma_{t-1} + \theta_2 \sigma_{t-2} \]  

(1)

where we - and not Barro! - made the assumption that $\eta_t$ is a second order moving average process.

Equation (1) is a standard Lucas supply function. The new hypothesis, that the $\eta_t$ is serially correlated, can be justified by saying that production presents scale-adjusting caused lags, etc.

The demand side of the economy is given by

\[ y_t = \beta_0 + \beta_1 t + \beta_2 W_t + \beta_3 P_t + \mu_t \] 

(2)

where $\mu_t$ is a white-noise process.
We are assuming instant market clearing, so that demand equals supply always.

Some comments ought to be made about equations (1) and (2).

First, we do not estimate the coefficients $\alpha_0$, $\alpha_1$, $\beta_0$, $\beta_1$. As said before, we are not interested in explaining the trend, but only the fluctuations away from it. Of course, we are assuming a very particular form of trend, a linear trend; we do not deny that this is perhaps a flaw in our model.

Second, one should see equation (2) as resulting from an idea similar to that of Keynes' consumption function. In this case we should use $n_t W_t$ instead of $W_t$, where $n_t$ is some measure of the quantity of people employed. We do not do so and since one has positive indication that $n_t$ is growing with $t$, we expect $\beta_2$ to be a biased measure of the nominal wage effect.

Third, we said that $n_t$ is a second-order moving average process while $v_t$ is a white noise. The ideas here are two: we allow demand to adjust faster than supply and we use these constraints to help identify the model.

Fourth, one may find it strange that we do not include a cash balance effect in (2). We do this because we want to study the power of prediction and the good fitness of our model when ignoring this type of effect.

Next, we assume that the economy follows the following wage contract rule:

$$W_t = \delta_0 + \delta_1 t + \delta_2 (y_t - E_{t+1} y_t) P_t + \varepsilon_t$$  \hspace{1cm} (3),$$

where again we will estimate neither $\delta_0$ nor $\delta_1$. We see (3) as a contract equation subject to the random errors $\varepsilon_t$. 
3. SOLVING THE MODEL

Our task now is to take the equations we have, put the necessary constraints that have not been put, and solve our rational-expectations model.

We shall first introduce the following notation:

\[ a_k = E_{t-1} y_{t+k}, \quad b_k = E_{t-1} p_{t+k}, \quad c_k = E_{t-1} w_{t+k}. \]

And, second, for the sake of clarity we repeat here what we have, renumbering the equations with letters:

\[ a_k = \alpha_0 + \alpha_1 t + \alpha_2 (p_t - E_t p_{t+1}) + \eta_t \]  

(A)

\[ \eta_t = \sigma_t + \theta_1 \sigma_{t-1} + \theta_2 \sigma_{t-2} \]  

(B)

\[ y_t = \beta_0 + \beta_1 t + \beta_2 w_t + \beta_3 p_t + \nu_t \]  

(C)

\[ w_t = \delta_0 + \delta_1 t + \delta_2 (y_t - E_{t-1} y_t) + \phi_t \]  

(D)

where \( \sigma_t, \nu_t, \phi_t \) are all white-noise.

Using rational expectations, we have for \( K > 2 \) that

\[ a_k = \alpha_0 + \alpha_1 t + \alpha_2 (b_k - b_{k+1}); \]

\[ a_k = \beta_0 + \beta_1 t + \beta_2 c_k + \beta_3 b_k; \]

\[ c_k = \delta_0 + \delta_1 t + b_k. \]

Solving for \( b_k \) we have that
\[ b_k = h_0 + h_1 t + C \lambda^k \]

where

\[ h_0 = \frac{\alpha_0 - \beta_0 - \beta_2 \delta_0}{\beta_2 + \beta_3} \]
\[ h_1 = \frac{\alpha_1 - \beta_1 - \beta_2 \delta}{\beta_2 + \beta_3} \quad \text{and} \]
\[ \lambda = \frac{\alpha_2 - (\beta_2 + \beta_3)}{\alpha_2} \]

We shall assume that the constant \( C \) is equal to zero. Indeed, in our estimations the value of \( \lambda \) is greater than 1 and hence if \( C \neq 0 \) we would have \( b_k \) exploding to \( \pm \infty \) when \( k \) grows indefinitely.

In this case, for \( k = 1 \) we have:

\[ a_1 = \alpha_0 + \alpha_1 t + \alpha_2 (b_1 - b_2) + \beta_2 \sigma_{t-1} \]
\[ a_1 = \beta_0 + \beta_1 t + \beta_2 \sigma_1 + \beta_3 b_1 \]
\[ c_1 = \delta_0 + \delta_1 t + b_1 \]

\( \implies \)
\[ b_1 = h_0 + h_1 t + h_2 \sigma_{t-2} \]

where \( h_3 = \frac{\beta_2}{\beta_2 + \beta_3 - \alpha_2} \)

Finally, for \( k = 0 \) we repeat the process and conclude that:

\[ b_0 = h_0 + h_1 t + h_2 \sigma_{t-1} + h_3 \sigma_{t-2} \]

where

\[ h_2 = \frac{1}{\beta_2 + \beta_3 - \alpha_2} \left[ \theta_1 - \alpha_2 h_3 \right] \]
We should point out that what we are obtaining is that the only mechanism of altering expectations of future inflations is by the transmission of errors and/or altering the coefficients $h_1$, $h_2$, $h_3$.

4. ESTIMATING THE MODEL

Using the lag notation we can write the resulting system as

$$A(L)z_t = B(L)e_t + h_t$$

where $A(L) = A_0$, $B(L) = B_0 + B_1L + B_2L^2$,

$$z_t' = (y_t, Z_t, W_t), e_t' = (\sigma_t, \mu_t, \phi_t)$$

and

$$h_t = C_1 + C_2t,$$ $C_1$ and $C_2$ being constant vectors.

The matrices $A_0$, $B_0$, $B_1$, $B_2$ are given by:

$$A_0 = \begin{bmatrix}
1 & -\alpha_2 & 0 \\
1 & 1 & -\beta_2 \\
-\delta_2 & 1 & 1
\end{bmatrix}$$

$$B_0 = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}$$
Now, instead of estimating the vector ARIMA we will estimate
\[
\begin{align*}
A(L) (z_t - z_{t-1} - \sum_{i=1}^{T} \frac{z_{t-i} - z_{t-1-i}}{T}) &= B(L) (I-L)e_t
\end{align*}
\]
which is a vector ARIMA for which I will use the Varma program. Notice that we are using \( \sum_{i=1}^{T} \frac{z_{t-i} - z_{t-1-i}}{T} \) as a proxy for \( C_1 \), which is clearly justified by the strong law of large numbers. Observe that
\[
\sum_{i=2}^{T} \frac{z_{t-i} - z_{t-1-i}}{T} = \frac{z_T - z_1}{T}.
\]

The implication of the last paragraph is that we will only have estimates for \( \alpha_2, \beta_2, \beta_3, \theta_1, \theta_2, \) and \( \delta_2 \).

We point out that for the second period we encountered problems with having Fisher's information matrix \( R(\theta_0) \) singular and had to proceed as described in appendix I. There we showed the need to use the estimators derived from the problem
\[
\begin{align*}
\max L(x, \theta) \\
s.t. \Pi_k \theta = 0
\end{align*}
\]
where \( L \) is the log-likelihood and \( \Pi_k \) is the projection on the kernel of \( R(\Theta_0) \).

The results are:

**FIRST PERIOD**
(1889-1914)

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimates</th>
<th>Asymptotic T</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_2 )</td>
<td>0.467</td>
<td>2.98</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-1.32</td>
<td>-10</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>-2.67</td>
<td>-8</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>0.52</td>
<td>2.68</td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>-0.15</td>
<td>-0.98</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>2.33</td>
<td>5.05</td>
</tr>
</tbody>
</table>

**SECOND PERIOD** *
(1909-1940)

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimates</th>
<th>Asymptotic T</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_2 )</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>5.905</td>
<td>2.02</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>1.016</td>
<td>5.97</td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>0.536</td>
<td>3.86</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

*We use the constraint \( \alpha_2 = \beta_3 = \delta_2 = 0 \) as a proxy for \( \Pi_k \Theta = 0 \). This was done by taking \( R(\Theta_0) \) in the free estimation and computing its eigenvalues. We found that the eigenvalues corresponding to \( \alpha_2, \beta_3 \) and \( \delta_2 \) were much smaller than the others.
We notice that for all periods $\alpha_2$ is greater than or equal to zero, in the first and last period supply falls with expected inflation. $\alpha_2 = 0$ in the second period was imposed in the estimation as a means of ruling out the non-consistency of the estimators obtained in the unconstrained estimation.

In the first period, $\beta_2$ has the sign opposite to what we expected. One possible explanation is that during that period the U.S. was receiving a huge mass of skilled labour from Europe at practically zero cost. If this exogenous growth of $n$ was negatively correlated with $W_t$, then using $n_t \cdot W_t$ in place of $W_t$ would give

$$\frac{\partial y_t}{\partial W_t} = \beta_2 (n_t + W_t) \frac{\partial n_t}{\partial W_t}$$

and this number could be negative.

We did not find any good explanation for the behaviour of $\delta_2$.

Empirically the model was very unsatisfactory for the third period, as shown by the correlation table between actual and predicted values below.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimates</th>
<th>Asymptotic T</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_2$</td>
<td>4.42</td>
<td>2.70</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>3.12</td>
<td>2.41</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-3.99</td>
<td>-3.23</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>0.83</td>
<td>6.16</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.07</td>
<td>0.97</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>-2.44</td>
<td>-8.00</td>
</tr>
<tr>
<td>VARIABLE</td>
<td>1st</td>
<td>2nd</td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>$Y_t$</td>
<td>0.5462</td>
<td>0.4718</td>
</tr>
<tr>
<td>$W_t$</td>
<td>0.3221</td>
<td>0.1979</td>
</tr>
<tr>
<td>$P_t$</td>
<td>0.2714</td>
<td>0.2628</td>
</tr>
</tbody>
</table>

Also, for the third period we faced the unpleasant feature of having $e_t$ autocorrelated.

5. A NOTE ON THE MONEY MARKET

Up to now we have not dealt with the money market, which may be one of the reasons that we obtained such bad results for the third period.

Suppose that the equilibrium of the monetary market is given by

$$M_t = y_0 + y_1^t + y_2^{M_{t-1}} + y_3^{M_{t-2}} + e_t$$  \hspace{1cm} (3)

where $e_t$ is a white noise. This equation presents the following fit for the periods considered:
Table I
FIRST PERIOD
(1889–1914)*

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>OLS-estimates</th>
<th>T-test for $H_0 = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>0.4208</td>
<td>2.169</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.0297</td>
<td>4.620</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.2999</td>
<td>-1.406</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.0174</td>
<td>1.918</td>
</tr>
</tbody>
</table>

$R^2 = 0.9908$  
$\bar{R}^2 = 0.9893$  
$F = 679.592$

Table II
SECOND PERIOD
(1909–1940)

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>OLS-estimates</th>
<th>T-test for $H_0 = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>0.381</td>
<td>2.095</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.0043</td>
<td>1.471</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>1.393</td>
<td>8.097</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-0.512</td>
<td>-3.000</td>
</tr>
</tbody>
</table>

$R^2 = 0.9806$  
$\bar{R}^2 = 0.9783$  
$F = 437.34$

* Since we have insufficient observations, we overlap the 1st and 2nd periods.

** $h_2$ is the correspondent of Durbin's h test when one has two lags.
Table III
THIRD PERIOD
(1959-1982)***

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>OLS-estimates</th>
<th>T-test for ( H_0 = 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 )</td>
<td>0.745</td>
<td>1.868</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>0.010</td>
<td>2.172</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>0.826</td>
<td>3.397</td>
</tr>
<tr>
<td>( \gamma_3 )</td>
<td>0.023</td>
<td>0.112</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.9988 \quad \text{DW} = 1.973 \]
\[ \frac{R^2}{R} = 0.9986 \quad h_2 = 0.973 \]
\[ F = 4959.38 \]

For all three periods the fit is very good and we point out that \( \lambda^2 - \gamma_2 \lambda - \gamma_3 = 0 \) always has different roots \( \lambda_1, \lambda_2 \), both lying inside the unit circle. In this case, it is easy to verify that (3) can be inverted into:

\[
\mu_t = \frac{\lambda_1}{\lambda_1 - \lambda_2} \sum_{i=0}^{\infty} \frac{\lambda_1^i}{\lambda_1 - \lambda_2} \varepsilon_{t-i} + \frac{\lambda_2}{\lambda_2 - \lambda_1} \sum_{i=0}^{\infty} \frac{\lambda_2^i}{\lambda_2 - \lambda_1} \varepsilon_{t-i} + \frac{\lambda_1}{\lambda_1 - \lambda_2} \left( \frac{\gamma_0 + \gamma_1 t}{1 - \lambda_1} + \frac{\lambda_1 \gamma_1}{(1 - \lambda_1)^2} \right) + \\
+ \frac{\lambda_2}{\lambda_2 - \lambda_1} \left( \frac{\gamma_0 + \gamma_1 t}{1 - \lambda_2} + \frac{\lambda_2 \gamma_1}{(1 - \lambda_2)^2} \right) \quad (4)
\]

*** When we initially did this regression, we had no data for 1953-1958.
One immediate consequence of accepting (4) is that the unexpected component of \( \mu_t \) is a white-noise. Another is that if we wrongly omit \( \mu_t \) from (2) then the correlation imposed by the first two terms of (4) will transfer to \( \mu_t \), giving the wrong impression that the \( \mu_t \)'s are correlated. The latter suggests that we might change our model by changing the hypothesis that \( \mu_t \) is a white noise. We shall assume that \( \mu_t \) is a first-order moving average process:

\[
\mu_t = f_t + \rho f_{t-1}.
\]

In this case, the model we obtain coincides with the old one, except that

\[
B_1 = \begin{bmatrix}
0 - \alpha_2 h_2 & -\frac{\alpha_2 \rho}{\beta_2 + \beta_3 - \alpha_2} & 0 \\
0 & \rho & 0 \\
-\alpha_2 h_2 & 0 & -\frac{\rho}{\beta_2 + \beta_3 - \alpha_2}
\end{bmatrix}
\]

It is a well known fact that increasing the order of an ARMA process increases the good fitness of the model, but that this procedure is misleading is also well known. Looking at the correlation between actual and predicted (one step ahead) values we obtain the following table:

| VARIABLE | PERIOD | | |
|----------|--------|---|---|---|
|          | 1st    | 2nd | 3rd |
| \( Y_t \) | 0.48086 | 0.7414 | 0.37463 |
| \( W_t \) | 0.40696 | 0.33564 | 0.84990 |
| \( P_t \) | 0.36024 | 0.58825 | 0.66818 |
The only significant change is then in the third period. For this period the estimate of the coefficients was then

**THIRD PERIOD**
**(1953-1982)**

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimates</th>
<th>Asymptotic T</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_2$</td>
<td>2.12</td>
<td>2.08</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.61</td>
<td>3.73</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.92</td>
<td>4.12</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>0.74</td>
<td>1.99</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.12</td>
<td>0.70</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>-4.65</td>
<td>-13.00</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.31</td>
<td>-3.39</td>
</tr>
</tbody>
</table>

6. **FINAL REMARKS**

By using theorem in the appendix, one can establish a test for comparing the coefficients across periods. We did this and the conclusion is that there is no qualitative difference between the first and second periods. So that $\alpha_2 = 0$, etc., should not be seen as very serious. We point out that this test has low power because of the number of variables minus the number of constrains with respect to the number of observations. Also, since we do not know the right order of moving average process involved, we are doomed to have a biased test.
For all three periods we have stability towards shocks on $e_t$.

Finally, the model assumes a very neutral government, one which has an almost blind monetary policy. This is far from reasonable and we think that it explains the low power of prediction presented, which is obviously true for the third period. Also, for the last period, there is something more going on, which is not captured by the model: there is a change of the correlation pattern between $D W_t$ and $D P_t$ compared with the pre-war period.
One of the most used theorems in Econometrics is that which says that maximum likelihood estimators (m.l.e.), under certain conditions, converge in probability to the true parameter of the distribution.

Our aim in this appendix is: first, to state the above theorem and prove it in its most general form; second, we want to know happens when some of the conditions used in the theorem are not true we will be most interested in what happens when the information matrix defined below is singular.

We start by a lemma that will prove itself useful:

**Lemma 1:** Given an mxm positive semi-definite matrix $A$ we can always find a positive definite matrix $A^+$ such that for every $v \in \text{Ker}A$ we have $Av = A^+v$.

**Proof:** Take $\Pi_k$ the projection into the kernel of $A$ and define $A^+ = A(I - \Pi_k) + \Pi_k$.

Then, if $v \in \text{Ker}A$, by definition, $\Pi_k v = 0 \Rightarrow A^+v = Av$. On the other hand, $v'A^+v = v'A(I-\Pi_k)v + v'\Pi_kv = v'(I-\Pi_k)'A(I-\Pi_k)v + v'\Pi_kv > 0$ if $v \neq 0 \Rightarrow A^+$ is positive definite.
With this theorem we can state the famous Carmer-Rao inequality in its most generalized form.

**Theorem 1: (Cramer-Rao inequality)**

Suppose $T$ is an unbiased estimator of $\theta$, that $L(.,\theta)$ is integrable and that $L(x,\theta)$ is twice differentiable in $\theta$ almost all $x$.

Then, if $R(\theta) = E \left( -\frac{\partial^2 \log f(x_i, \theta)}{\partial \theta^1 \partial \theta^1} \right)$ is defined, we have

$$\text{Cov} T = 2 (R(\theta)^+)^{-1} + (R(\theta)^+)^{-1} R(\theta) (R(\theta)^+)^{-1}$$

is positive semi-definite.

The proof of this theorem can be obtained with little modification from the one in Chow, page 23.

Notice that what (A-1) is actually doing is setting a lower bound for $\text{Cov} T$. We shall say that $T$ is asymptotically efficient when

$$\text{Cov} T = 2 (R(\theta)^+)^{-1} - (R(\theta)^+)^{-1} (R(\theta) (R(\theta)^+)^{-1}$$

A characterization of $A^+$ given $A$ would be most desirable here, however we shall only indicate to the reader that the way to characterize $A^+$ passes through using Jordan's canonical form (see Hoffman-Kunze, chapter 7). Indeed, this yields another proof of lemma 1. We shall see, however, that for our applications we will not need it.

**Definition:** Let $X_1, X_2, \ldots, X_n$ be $n$ independent identically distributed (i.i.d.) random variables having density function $f(x, \theta)$, where $\theta$ is a finite-dimensional vector of parameters. We define the $n$th log-likelihood function as $L(x_1, \ldots, x_n, \theta) = \sum_{i=1}^{n} \log f(x_i, \theta)$. 

$$L(x_1, \ldots, x_n, \theta) = \sum_{i=1}^{n} \log f(x_i, \theta).$$
Given a realization of $\hat{x}_1, \ldots, \hat{x}_n$ a maximum likelihood estimator \( \theta \) is a point of maximum of $L(\hat{x}_1, \ldots, \hat{x}_n, \theta)$.

We shall need the following two theorems from Classical Probability Theory:

**Theorem 2**: (Strong Law of Large Numbers)

Let $X_n$ be a sequence of independent and identically distributed random variables. Then we have

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i \to E(X) \text{ a.e.}
$$

**Lemma 2**: Suppose $\mathbb{R}$ is compact and convex and that, \( E \log f(x,\theta) \) is a strictly convex and continuous function in $\theta$. Let $\hat{\theta}_n$ be any sequence of m.l.e.’s, then $\hat{\theta}_n \xrightarrow{a.e.} \theta$ where $\theta$ maximum of $E \log f(x, \theta)$.

**Proof**: Suppose $\theta_n \xrightarrow{a.e.} \hat{\theta}$, then $\hat{\theta} = \theta$.

Indeed,

$$
\frac{1}{n} L(\hat{x}_1, \ldots, \hat{x}_n, \hat{\theta}_n) > \frac{1}{n} L(\hat{x}_1, \ldots, \hat{x}_n, \theta)
$$

and using the Law of Large Numbers we get

$$
E \log f(x, \theta) \xrightarrow{a.e.} E \log f(x, \theta)
$$

and since $\theta$ is the unique maximum we get $\theta = \hat{\theta}$ a.e.

**Theorem 3**: (Central Limit Theorem)

Let $X_n$ be a sequence of i.i.d. random variables with finite non-singular covariance matrix Cov $X_1$, then
\[
\left(\text{Cov } X_1\right)^{-1/2} \frac{\sum_{i=1}^{n} (X_i - \bar{X})}{\sqrt{n}} \overset{D}{\to} N(0, I),
\]

where \(I\) is the identity matrix with the same number of rows of \(\text{Cov } X_1\).

We are not going to prove either of these two theorems. Their proof can be found in Chung's book \cite{3}. Notice that in the case \(\text{Cov } X_1\) is not a scalar the notation \(\left(\text{Cov } X_1\right)^{-1/2}\) really means the matrix \(H\) such \(H' H = \left(\text{Cov } X_1\right)^{-1}\).

Since \(R\) is compact, given a sequence \(\theta\) we can always extract a converging subsequence \(\tilde{\theta}_n'\). By the paragraph above the limit of \(\tilde{\theta}_n'\) is equal to \(\tilde{\theta}\) a.e.. This implies that \(\tilde{\theta}_n\) converges a.e. to \(\tilde{\theta}\).

**Theorem 4**: under the conditions of lemma 2 if we suppose \(\theta_0 \in \text{int } \Omega\), that \(L\) is twice differentiable and that \(R(\theta_0)\) is defined and positive definite, we have that the \(\theta_n\) are asymptotically normal.

**Proof**: The m.l.e. \(\theta_n\) is a solution to \(\frac{\partial L}{\partial \theta} (x, \theta_n) = 0\), use this to write

\[
\theta = \frac{\partial L}{\partial \theta} (x, \theta_n) = \frac{\partial L}{\partial \theta} (x, \theta_0) + \frac{\partial^2 L}{\partial \theta^2} (x, \theta_0) (\theta_n - \theta_0) + o(\theta_n - \theta_0) \quad (A-2)
\]

where \(\lim_{\tilde{\theta}_n \to \theta_0} \frac{\sigma(\theta_n - \theta_0)}{||\theta_n - \theta_0||} = 0\). Then multiply this equation by \(R(\theta_0)^{-1}\) and take the limit when \(n \to \infty\). By the strong law

\[
\frac{1}{n}(\theta_0)^{-1} \left( \frac{\partial^2 L}{\partial \theta^2} (x, \theta_0) \right) \overset{\text{a.e.}}{\to} - I
\]
and by the Central Limit theorem \( \text{Cov} \left( \frac{3L}{\partial \Theta} \right) = R(\Theta_0) \), we have that

\[
\frac{1}{\sqrt{n}} R(\Theta_0)^{-1} \frac{3L}{\partial \Theta} (x, \Theta_0) \xrightarrow{D} N(0, R(\Theta_0)^{-1}). \quad \text{And so}
\]

\[
\sqrt{n} (\Theta_n - \Theta_0) \xrightarrow{D} N(0, R(\Theta_0)^{-1}). \quad \text{QED}
\]

Nothing up to now indicates that \( \Theta_0 \) is a true parameter of the distribution. Indeed, what is usually done is to take this hypothesis and give further justification. What happens, then, when we have no convergence assured? The obvious answer has to come from the analysis of lemma 2 above. We will only look at the condition saying that \( \text{Elog}(x_i, \Theta) \) is strictly concave in \( \Theta \).

Now, if we want not only to avoid technical problems that will only complicate the proofs, but also preserve the spirit of the results, we will allow \( L \) to be twice differentiable. Suppose, then, that we solve

\[
\max \ L(x, \Theta) \\
\text{s.t.} \ \Pi_k \Theta = 0
\]

where here \( \Pi_k \) is the projection on the kernel of \( R(\Theta_0) \). It is easy to check that \( L \) is strictly concave outside the kernel of \( R(\Theta_0) \) and so we can mimic lemma 2 to find estimators \( \Theta_n \) with limit \( \Theta_0 \), the only difference being that now \( \Pi_k \Theta_0 = 0 \). Theorem 4 can also be copied yet we need to take some care: firstly, the first-order condition for maximization is now given by

\[
\frac{3L}{\partial \Theta} (x, \Theta_n) = \Pi_k \lambda \quad \text{(A-3)}
\]

\[
\Pi_k \Theta_n = 0
\]

where \( \lambda = (\lambda_1, \ldots, \lambda_m) \) is a vector of lagrange multipliers; secondly, because of this (A-2) is written
\[ \Pi_k^\lambda = -\frac{\partial L}{\partial \theta}(x, \theta_0) + \frac{\partial^2 L}{\partial \theta^2}(x, \theta_0) (\theta_n - \theta_0) + \sigma(\theta_n - \theta_0) \]

Multiplying this equality by \((\theta_n - \theta_0)'(R(\theta_0)^+)^{-1}\) we have

\[ 0 = (\theta_m - \theta_0)'[ (R(\theta_0)^+)^{-1} \frac{\partial L}{\partial \theta}(x, \theta_0) + \]

\[ + (R(\theta_0)^+)^{-1} \frac{\partial^2 L}{\partial \theta^2}(x, \theta_0) (\theta_m - \theta_0) + \sigma(\theta_m - \theta_0) ] \]

But this implies

\[ (R(\theta_0)^+)^{-1} \frac{\partial L}{\partial \theta}(x, \theta_0) + (R(\theta_0)^+)^{-1} \frac{\partial^2 L}{\partial \theta^2}(x, \theta_0) (\theta_m - \theta_0) \]

\[ + \sigma(\theta_m - \theta_0) \in \text{Ker } R(\theta_0) \]

\[ \Rightarrow (I - \Pi_k^\lambda) (R(\theta_0)^+)^{-1} \frac{\partial L}{\partial \theta}(x, \theta_0) + \]

\[ + (I - \Pi_k^\lambda) (R(\theta_0)^+)^{-1} \frac{\partial^2 L}{\partial \theta^2}(x, \theta_0) (\theta_n - \theta_0) + \]

\[ + (I - \Pi_k^\lambda) \sigma(\theta_n - \theta_0) = 0 \]

and, then, using the strong law, the Central Limit Theorem and the fact that \((R(\theta_0)^+)^{-1} R(\theta_0) = I - \Pi_k\) we conclude that:

\[ \sqrt{n} (I - \Pi_k^\lambda) (\theta_m - \theta_0) \xrightarrow{D} N(0, (I - \Pi_k^\lambda)(R(\theta_0)^+)^{-1} (I - \Pi_k^\lambda)') \]  \((\text{A-4})\)

The D-limit above is the maximum one can say when \(R(\theta_0)\) is singular using this kind of technique. Suppose that we want to test \(H_0 : \theta_{\text{true}} = \theta_0\) against \(H_1 : \theta_{\text{true}} = \theta_1\); using (A-4) above, what we actually obtain is a cylinder corresponding to the perpendicular translation of the confidence region in \(\text{Ker } R(\theta_0)\) obtained by testing the positive components of \((I - \Pi_k^\lambda)(\theta_n - \theta_1)\). This indicates that the power of (A-4) should not be very good.

The problem is then to find another test for which we would have good power. The answer was given by Haussmann ([4]), who ori-
Finally studied the following situation: suppose we have a model

\[ y = X\beta + \alpha + v \]  

(A-5)

and we want to study \( H_0 : \alpha = 0 \) against \( H_1 : \alpha \neq 0 \); a priori no information is given about the \( v \)'s, if the \( v \)'s are correlated with \([X \bar{X}]\), etc. this result was the following theorem:

**Theorem 5:** Suppose we have two estimators \( \theta_0, \theta_1 \), both consistent and asymptotically normally distributed with \( \theta_0 \) attaining the asymptotic Cramer-Rao bound. Let, also, \( m \) be the number of coordinates in each of these coordinates, \( q = (\theta_1 - \text{plim} \, \theta_1) - (\theta_0 - \text{plim} \, \theta_0) \) and \( M = V(\theta_1) - V(\theta_0) \), where \( V \) indicates the covariance matrix, then

\[ nq'(M^+)^{-1}M(M^+)^{-1}q \xrightarrow{D} \chi^2_k \]

where \( k = m - \text{dim} \, (\ker M) \).

To complete this appendix we extend Durbin's h-test, see Chow [2] page 85, to the case where one has \( s \) legged dependent variables on the right hand side of a regression \( y = X\beta + \mu \) and want to test if \( \mu \) is autocorrelated. The basic reference is Breush and Pagan [5]. There the authors, by using Lagrange estimators, conclude that if one has the model

\[ y = X\beta + \mu, \ X \text{ is } NxK \]

\[ \mu_t = \rho \mu_{t-1} + e_t \]

where \( e_t \sim N(0, \sigma^2 I) \), one has that the null hypothesis \( H_0 : \rho = 0 \), can be tested by using the statistic

\[ T = N \rho^2 \left[ 1 - \sigma^{-2} \left( \frac{1}{\mu-1} X \right) (N^{-1}X'X)^{-1} (N^{-1}X'X) \frac{1}{\mu-1} \right]^{-1} \]
where: \( \mu' = (\hat{\mu}_N', \ldots, \hat{\mu}_{s+2})' \), \( \hat{\mu}_{-1} = (\hat{\mu}_{N-1}', \ldots, \hat{\mu}_{s+1})' \) the \( - \) indicating that \( \mu \) is estimated under \( H_0 = 0 \) (OLS); and \( x_1^2 = (\hat{\mu}', \hat{\mu})^{-1} \hat{\mu}_{-1} \hat{\mu}' \).

They prove that \( T \overset{D}{\rightarrow} \chi_1^2 \).

Our task then is just to calculate the plim of \( \hat{\sigma}^{-2} N^{-1} \hat{\mu}_{-1} X \) under the s lags condition. We partition \( X \) into two matrices \( X = (X_e, X_{ne}) \), \( X_e \) containing all the lagged values:

\[
X_e = \begin{bmatrix}
Y_{N-1} & Y_{N-2} & \cdots & Y_{N-s} \\
Y_{N-2} & Y_{N-3} & \cdots & Y_{N-s-1} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{s+1} & \cdots & Y_{1} & Y_{-1} & Y_{-2} & \cdots & Y_{-s}
\end{bmatrix}
\]

Now given that \( \mu_{-1} = M Y_{-1} \), where \( M = I - X(X'X)^{-1}X' \), we use the Weak Law of Large Numbers to get

\[
\frac{Y_{-1}' \mu_{-1}}{\hat{\sigma}^{-2} N} \overset{P}{\rightarrow} 1
\]

\[
\frac{Y_{-1}' \mu_{-2}}{\hat{\sigma}^{-2} N} \overset{P}{\rightarrow} \beta_1
\]

\[
\frac{Y_{-1}' \mu_{-s}}{\hat{\sigma}^{-2} N} \overset{P}{\rightarrow} \beta_{s-1}
\]

and

\[
\frac{Y_{-1}' M X_{ne}}{\hat{\sigma}^{-2} N} \overset{P}{\rightarrow} 0
\]

\[
= \hat{\sigma}^{-2} N^{-1} \hat{\mu}_{-1} X \overset{P}{\rightarrow} (1, \beta_1, \ldots, \beta_{s-1})
\]
where $\beta_1, \ldots, \beta_s$ are the coefficients of $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-s}$.

For the case $s = 2$ we have

$$T = \frac{N r_1^2}{\left[1 + \text{Cov} \beta_1 + 2 \beta_1 \cdot \text{Cov}(\beta_1, \beta_2) + \beta_1^2 \cdot \text{Cov} \beta_2 \right]}$$

(in the text $h_2 = \sqrt{T}$).
BIBLIOGRAPHY


