# Lecture 7: Processes with Deterministic Trends*

## 1 Introduction

Recall that a process is covariance stationary if it has constant expectation, finite variance, and its antocovariance functions do not depend on time. In this lecture, we will introduce one class of processes that are nonstationary – processes with deterministic trend. In the next lecture, we will introduce another type – processes with stochastic trend.

We have been familiar with a stationary ARMA process,

$$\tilde{x}_t = \psi(L)u_t.$$

Now consider an ARMA process with a drift,

$$x_t = \delta t + \tilde{x}_t = \delta t + \psi(L)u_t. \tag{1}$$

Now the expectation of $x_t$ is $\delta t$, which is a function of time, so this process is nonstationary. We can decompose the process $x_t$ into two components: a trend component ($\delta t$) and a stationary component ($\tilde{x}_t$). If $\delta$ is known, then we can *detrend* $x_t$, i.e., $x_t - \delta t$ to get $\tilde{x}_t$, which is a stationary process, so the process $\{x_t\}$ is said to be *trend stationary*.

The $k$-period ahead forecasting of $x$ is

$$\begin{aligned}
& E_t(x_{t+k}) \\
= \; & E_t\left(\delta(t+k) + \psi(L)u_{t+k}\right) \\
= \; & \delta(t+k) + E_t(u_{t+k} + \psi_1 u_{t+k-1} + \ldots + \psi_k u_t + \psi_{k+1}u_{t-1} + \ldots + \psi_{t+k}u_0) \\
= \; & \delta(t+k) + \psi_k u_t + \psi_{k+1}u_{t-1} + \ldots + \psi_{t+k}
\end{aligned}$$

And the forecasting error is:

$$\begin{aligned}
& E_t(x_{t+k} - E_t(x_{t+k}))^2 \\
= \; & E_t(u_{t+k} + \psi_1 u_{t+k-1} + \ldots + \psi_{k-1}u_{t+1})^2 \\
= \; & (1 + \psi_1^2 + \psi_2^2 + \ldots + \psi_{k-1}^2)\sigma^2
\end{aligned}$$

Note since $\tilde{x}_t = \psi(L)\epsilon_t$ is a stationary process, as $k \to \infty$, the forecasting error converges to the unconditional variance of $\tilde{x}_t$, which is bounded. This is a very important difference between processes with deterministic trend and those with stochastic trend. Another feature of a trend stationary process is that given a shock at time $t$, its effects on the level of $\tilde{x}$, hence $x$ eventually dies off as in a stationary process. This is another difference compared to a unit root process. We will discuss more on this in next lecture.

Figure 1 plots a simulated path of (1), where $\delta = 1$, $u_t \sim N(0, 1)$ and $\psi(L) = 1 + 0.5L$.

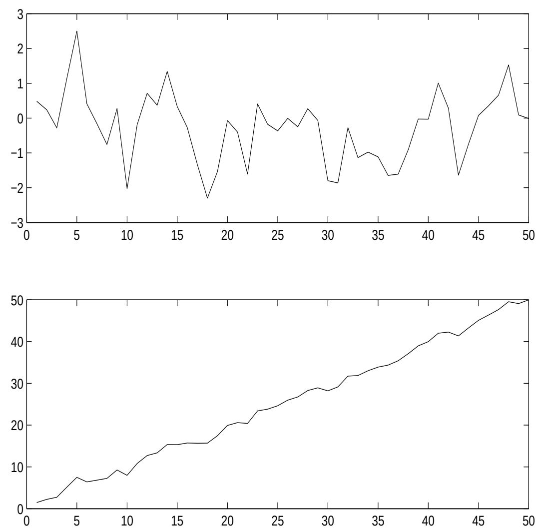


Figure 1: Simulated MA(2) Process with Deterministic Trend

# 2 Estimation and Inference

## 2.1 OLS estimation of the simple time trend model

Consider the process

$$y_t = \alpha + \delta t + u_t = x_t\beta + u_t, \tag{2}$$

where $\beta' = [\alpha, \delta]$, $x_t' = [1, t]$, and $u_t \sim i.i.d.N(0, \sigma^2)$. We can use MLE to estimate the parameters $\beta$, and the MLE estimator is equivalent to the OLS estimator. So we will only discuss OLS estimation, which is applicable to a more general class of errors. In our following analysis, we assume $u_t \sim i.i.d(0, \sigma^2)$ and $E(u_t^4) < \infty$.

The OLS estimate of $\beta$ is

$$\begin{aligned} \hat{\beta}_n &= \left[\sum_{t=1}^n x_t x_t'\right]^{-1} \left[\sum_{t=1}^n x_t y_t\right] & (3) \\ &= \beta_0 + \left[\sum_{t=1}^n x_t x_t'\right]^{-1} \left[\sum_{t=1}^n x_t u_t\right] & (4) \end{aligned}$$

Since $x_t$ is deterministic, take expectation of $\hat{\beta}_n$, we have $E(\hat{\beta}_n) = \beta_0$. So $\hat{\beta}_n$ is an unbiased estimator for $\beta_0$. It can be also shown that they are consistent (converges to the true value). So far what we got are just the same as what we got with a stationary processes. However, although $\hat{\beta}_n = (\hat{\alpha}_n, \hat{\delta}_n)'$ converges to the true parameter $\beta_0 = (\hat{\alpha}_0, \hat{\delta}_0)'$, it turns out that its two components $\hat{\alpha}_n$ and $\hat{\delta}_n$ converge at different rates!

To see this, note that $x_t x_t'$ is a 2 by 2 matrix,

$$\sum_{t=1}^n x_t x_t' = \begin{bmatrix} n & \sum_{t=1}^n t \\ \sum_{t=1}^n t & \sum_{t=1}^n t^2 \end{bmatrix}.$$

Some simple math gives

$$\sum_{t=1}^{n} t = n(n+1)/2 = O(n^2),$$
$$\sum_{t=1}^{n} t^2 = n(n+1)(2n+1)/6 = O(n^3),$$

or

$$\frac{1}{n^2}\sum_{t=1}^{n} t \to \frac{1}{2} \quad \frac{1}{n^3}\sum_{t=1}^{n} t^2 \to \frac{1}{3},$$

More generally, we have

$$\frac{1}{n^{r+1}}\sum_{t=1}^{n} t^r \to \frac{1}{r+1}.$$

So the elements of matrix of $X_n'X_n$ diverges at different rate. To obtain a convergent matrix, we have to divide it by $n^3$ (the largest divergent rate),

$$n^{-3}\sum_{t=1}^{n} x_t x_t' = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}.$$

Unfortunately, this limiting matrix is singular and cannot be inverted. It turns out that to obtain a nondegenerate limiting distributions, $\hat{\alpha}_n$ need to be rescaled by $n^{1/2}$ and $\hat{\delta}_n$ need to be rescaled by $n^{3/2}$. Therefore, to get a proper limit of $\hat{\beta}_n$, we need to normalize it with a matrix

$$H_n = \begin{bmatrix} n^{1/2} & 0 \\ 0 & n^{3/2} \end{bmatrix}.$$

Now premultiply $\hat{\beta}_n$ by $H$.

$$\begin{aligned} H_n(\hat{\beta}_n - \beta_0) &= \begin{bmatrix} n^{1/2}(\hat{\alpha}_n - \alpha_0) \\ n^{3/2}(\hat{\delta}_n - \delta_0) \end{bmatrix} \\ &= H_n \left[\sum_{t=1}^{n} x_t x_t'\right]^{-1} H_n H_n^{-1} \left[\sum_{t=1}^{n} x_t u_t\right] \\ &= \left[H_n^{-1}\left(\sum_{t=1}^{n} x_t x_t'\right) H_n^{-1}\right]^{-1} \left[H_n^{-1}\left(\sum_{t=1}^{n} x_t u_t\right)\right] \end{aligned}$$

We first drive the limit for the matrix $H_n^{-1}X_n'X_nH_n$.

$$H_n^{-1}\left(\sum_{t=1}^{n} x_t x_t'\right) H_n^{-1} = \begin{bmatrix} n^{1/2} & 0 \\ 0 & n^{3/2} \end{bmatrix}^{-1} \begin{bmatrix} n & \sum_{t=1}^{n} t \\ \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 \end{bmatrix} \begin{bmatrix} n^{1/2} & 0 \\ 0 & n^{3/2} \end{bmatrix}^{-1} \to \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}$$

as $n \to \infty$. We will use $Q$ to denote this matrix,

$$Q = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}.$$

Next, we need to derive the asymptotic distribution for $H_n^{-1}\left(\sum_{t=1}^n x_t u_t\right)$,

$$H_n^{-1}\left(\sum_{t=1}^n x_t u_t\right) = \begin{bmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix} \begin{bmatrix} \sum_{t=1}^n u_t \\ \sum_{t=1}^n t u_t \end{bmatrix} = \begin{bmatrix} n^{-1/2}\sum_{t=1}^n u_t \\ n^{-1/2}\sum_{t=1}^n (t/n) u_t \end{bmatrix}$$

We will show that this vector is asymptotically normal with mean zero and covariance matrix $\sigma^2 Q$. First consider the term $n^{-1/2}\sum_{t=1}^n u_t$, applying the classical central limit theorem directly, we have

$$n^{-1/2}\sum_{t=1}^n u_t \to N(o, \sigma^2).$$

Second, consider the term $n^{-1/2}\sum_{t=1}^n (t/n)u_t$. Now the series $\{(t/n)u_t\}$ is not i.i.d, but it is a martingale difference sequence, and we can apply the CLT for mds. To apply the CLT of mds, we need to show that the three conditions in proposition 15 in lecture note 4 (proposition 7.8 in Hamilton) are satisfied. First, $E[(t/n)u_t]^2 = (t^2/n^2)\sigma^2$, and $n^{-1}\sum_{t=1}^n (t^2/n^2)\sigma^2 \to \sigma^2/3 > 0$, so condition (a) is satisfied. Second, take $r = 4$, since $u_t$ has finite fourth moment by assumption, condition (b) is satisfied. Finally, we need to show that

$$n^{-1}\sum_{t=1}^n (t^2/n^2)u_t^2 \to \sigma^2/3.$$

Since we have that

$$n^{-1}\sum_{t=1}^n (t^2/n^2)\sigma^2 \to \sigma^2/3,$$

we just need to show that

$$n^{-1}\sum_{t=1}^n (t^2/n^2)(u_t^2 - \sigma^2) \to 0. \tag{5}$$

Note that the series $\{(t^2/n^2)(u_t^2 - \sigma^2)\}$ is a mds with variance

$$((t^4/n^4)E[(u_t^2 - \sigma^2)^2] = (t^4/n^4)[E(u_t^4) - \sigma^4] = (t^4/n^4)(\mu_4 - \sigma^4) < \infty,$$

So (5) holds by law of large numbers for mds. Now all the three conditions are satisfied, we can then apply the CLT for mds,

$$n^{-1/2}\sum_{t=1}^n (t/n)u_t \to N(0, \sigma^2/3).$$

The remaining task is to show that $\{n^{-1/2}\sum_{t=1}^n u_t\}$ and $\{n^{-1/2}\sum_{t=1}^n (t/n)u_t\}$ are asymptotically joint normal. To show they are jointly normal, it is suffice to show that any linear combination of these two series is asymptotically normal, i.e., to show that

$$n^{-1/2}\sum_{t=1}^n [\lambda_1 + \lambda_2(t/n)]u_t \to N(0, \Sigma).$$

Note that the series $\{\lambda_1 u_t + \lambda_2 (t/n) u_t\}$ is a mds with variance $\sigma^2[\lambda_1^2 + 2\lambda_1\lambda_2(t/n) + \lambda_2^2(t/n)^2]$ satisfying

$$\frac{1}{n}\sum_{t=1}^{n}\sigma^2[\lambda_1^2 + 2\lambda_1\lambda_2(t/n) + \lambda_2^2(t/n)^2] \to \sigma^2[\lambda_1^2 + 2\lambda_1\lambda_2(1/2) + \lambda_2^2(1/3)] = \sigma^2\boldsymbol{\lambda}'Q\boldsymbol{\lambda}$$

for $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$. Furthermore,

$$\frac{1}{n}\sum_{t=1}^{n}[\lambda_1 + \lambda_2(t/n)]^2\epsilon_t^2 \to \sigma^2\boldsymbol{\lambda}'Q\boldsymbol{\lambda}.$$

So we can apply CLT and have this linear combination of the two elements converge to a Gaussian distribution, this hence imply that this two elements are joint Gaussian.

$$\begin{bmatrix} n^{-1/2}\sum_{t=1}^{n} u_t \\ n^{-1/2}\sum_{t=1}^{n}(t/n)u_t \end{bmatrix} \to N(0, \sigma^2 Q).$$

Therefore, we got

$$H_n^{-1}\left(\sum_{t=1}^{n} x_t u_t\right) = \begin{bmatrix} n^{-1/2}\sum_{t=1}^{n} u_t \\ n^{-3/2}\sum_{t=1}^{n} t u_t \end{bmatrix} \to N(0, \sigma^2 Q^{-1} Q Q^{-1}) = N(0, \sigma^2 Q^{-1}).$$

We can summarize the results in

**Proposition 1** *Let $y_t$ be generated according to the simple deterministic time trend model (2) where $u_t \sim i.i.d.(0, \sigma^2)$ with finite fourth moment. Then*

$$\begin{bmatrix} n^{1/2}(\hat{\alpha}_n - \alpha) \\ n^{3/2}(\hat{\delta}_n - \delta) \end{bmatrix} \to N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}^{-1}\right).$$

Note that for the estimate of $\delta$, we not only have $\hat{\delta}_n \to_p \delta$, we also have $n(\hat{\delta}_n - \delta) \to_p 0$. In this case, the estimate $\hat{\delta}$ is said to be *superconsistent*.

## 2.2 Hypothesis testing for the simple time trend model

When the innovation term $u_t$ is Gaussian, and since in the simple trend model the regressors are deterministic, the OLS estimates $\hat{\alpha}_n$ and $\hat{\delta}_n$ are Gaussian and the usual OLS $t$ and $F$ tests have the exact small sample $t$ and $F$ distribution. In this section, we will consider the case when $u_t$ is non-Gaussian.

We first consider a test of the null hypothesis on $\alpha$, say, $\alpha = a$. Let $s_n^2$ is the OLS estimate of $\sigma^2$: $s_n^2 = \frac{1}{n-2}\sum_{t=1}^{n}\hat{u}_t^2$. Then the $t$ statistics is

$$\begin{aligned} t_n &= \frac{\hat{\alpha}_n - a}{\left\{s_n^2 \begin{bmatrix} 1 & 0 \end{bmatrix} (X_n' X_n)^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}^{1/2}} \\ &= \frac{\sqrt{n}(\hat{\alpha}_n - a)}{\left\{s_n^2 \begin{bmatrix} \sqrt{n} & 0 \end{bmatrix} (X_n' X_n)^{-1} \begin{bmatrix} \sqrt{n} \\ 0 \end{bmatrix}\right\}^{1/2}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\sqrt{n}(\hat{\alpha}_n - a)}{\left\{ s_n^2 \begin{bmatrix} 1 & 0 \end{bmatrix} H_n (X_n' X_n)^{-1} H_n \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}^{1/2}} \\
&\to \frac{\sqrt{n}(\hat{\alpha}_n - a)}{\left\{ \sigma^2 \begin{bmatrix} 1 & 0 \end{bmatrix} Q^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}^{1/2}}
\end{aligned}
$$

where we uses that $s_n^2 \to_p \sigma^2$,

$$
\begin{bmatrix} \sqrt{n} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} H_n \quad \text{and} \quad H_n (X_n' X_n)^{-1} H_n = [H_n^{-1} (X_n' X_n) H_n^{-1}]^{-1} \to Q^{-1}.
$$

Let $q_{11}$ denote the $(1, 1)$ element of $Q^{-1}$, then under the null hypothesis we know that $\sqrt{n}(\hat{\alpha}_n - a) \to N(0, \sigma^2 q_{11})$. So we can see that

$$
t_n \to \frac{\sqrt{n}(\hat{\alpha}_n - a)}{\sigma \sqrt{q_{11}}},
$$

is an asymptotically Gaussian variable divided by the square root of its variance, so it has a $N(0, 1)$ distribution.

Similarly, to test the null hypothesis $\hat{\delta}_n = b$, write

$$
\begin{aligned}
t_n &= \frac{\hat{\delta}_n - b}{\left\{ s_n^2 \begin{bmatrix} 0 & 1 \end{bmatrix} (X_n' X_n)^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}^{1/2}} \\
&= \frac{n^{3/2}(\hat{\delta}_n - b)}{\left\{ s_n^2 \begin{bmatrix} 0 & n^{3/2} \end{bmatrix} (X_n' X_n)^{-1} \begin{bmatrix} 0 \\ n^{3/2} \end{bmatrix} \right\}^{1/2}} \\
&= \frac{n^{3/2}(\hat{\delta}_n - b)}{\left\{ s_n^2 \begin{bmatrix} 0 & 1 \end{bmatrix} H_n (X_n' X_n)^{-1} H_n \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}^{1/2}} \\
&\to \frac{n^{3/2}(\hat{\delta}_n - b)}{\sigma \sqrt{q_{22}}},
\end{aligned}
$$

which is again asymptotically $N(0, 1)$.

We have just considered tests on either $\alpha$ or $\delta$. Now consider a test involving both $\alpha$ and $\delta$:

$$
H_0: \; r_1 \alpha + r_2 \delta = r.
$$

We will apply similar procedure as before, but $\alpha$ and $\delta$ have different convergent rate $n^{1/2}$ and $n^{3/2}$, which one shall we use to derive the asymptotics? It turns out (again) that the slower rate 'dominates'.

$$
t_n = \frac{(r_1 \hat{\alpha}_n + r_2 \hat{\delta}_n - r)}{\left\{ s_n^2 \begin{bmatrix} r_1 & r_2 \end{bmatrix} (X_n' X_n)^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \right\}^{1/2}}
$$

$$= \frac{\sqrt{n}(r_1\hat{\alpha}_n + r_2\hat{\delta}_n - r)}{\left\{ s_n^2 \sqrt{n} \begin{bmatrix} r_1 & r_2 \end{bmatrix} (X_n' X_n)^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \sqrt{n} \right\}^{1/2}}$$

$$= \frac{\sqrt{n}(r_1\hat{\alpha}_n + r_2\hat{\delta}_n - r)}{\left\{ s_n^2 \sqrt{n} \begin{bmatrix} r_1 & r_2 \end{bmatrix} H_n^{-1} H_n (X_n' X_n)^{-1} H_n H_n^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \sqrt{n} \right\}^{1/2}}$$

$$= \frac{\sqrt{n}(r_1\hat{\alpha}_n + r_2\hat{\delta}_n - r)}{\{ s_n^2 \mathbf{r}_n' [H_n (X_n' X_n)^{-1} H_n] \mathbf{r}_n \}^{1/2}},$$

where

$$\mathbf{r}_n \equiv H_n^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \sqrt{n} = \begin{bmatrix} r_1 \\ r_2/n \end{bmatrix} \to \begin{bmatrix} r_1 \\ 0 \end{bmatrix}.$$

Since $\hat{\delta}_n$ is superconsistent,

$$\sqrt{n}(r_1\hat{\alpha}_n + r_2\hat{\delta}_n - r) = \sqrt{n}(r_1\hat{\alpha}_n + r_2\delta - r) + o_p(1).$$

So

$$t_n \to_p \frac{\sqrt{n}(r_1\hat{\alpha}_n + r_2\hat{\delta}_n - r)}{\left\{ \sigma^2 \begin{bmatrix} r_1 & 0 \end{bmatrix} Q^{-1} \begin{bmatrix} r_1 \\ 0 \end{bmatrix} \right\}^{1/2}} = \frac{\sqrt{n}(r_1\hat{\alpha}_n + r_2\delta - r)}{(r_1^2 \sigma^2 q_{11})^{1/2}} + o_p(1).$$

Further, note that

$$\sqrt{n}(r_1\hat{\alpha}_n + r_2\delta - r) = \sqrt{n}[r_1(\hat{\alpha}_n - \alpha) + r_1\alpha + r_2\delta - r) = \sqrt{n}[r_1(\hat{\alpha}_n - \alpha)]$$

under the null hypothesis. Therefore, under the null

$$t_n \to_p \frac{\sqrt{n}[r_1(\hat{\alpha}_n - \alpha)]}{(r_1^2 \sigma^2 q_{11})^{1/2}} = \frac{\sqrt{n}(\hat{\alpha}_n - \alpha)}{\sigma \sqrt{q_{11}}}.$$

which is asymptotically $N(0,1)$. This example shows that a test involving a single restriction across parameters with different rates of convergence is dominated asymptotically by the parameters with the slowest rates of convergence.

Finally consider joint test of separate hypothesis about $\alpha$ and $\delta$,

$$H_0 : \begin{bmatrix} \alpha \\ \delta \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

or in vector form, $\beta = \mathbf{c}$. Then we could compute a Wald statistics

$$\begin{aligned} W_n &= (\hat{\beta}_n - \mathbf{c})' [s_n^2 (X_n' X_n)^{-1}]^{-1} (\hat{\beta}_n - \mathbf{c}) \\ &= (\hat{\beta}_n - \mathbf{c})' H_n [s_n^2 H_n (X_n' X_n)^{-1} H_n]^{-1} H) n (\hat{\beta}_n - \mathbf{c}) \\ &\to [H_n(\hat{\beta}_n - \mathbf{c})]' [\sigma^2 Q^{-1}]^{-1} [H_n(\hat{\beta}_n - \mathbf{c})]. \end{aligned}$$

Then we have

$$W_n \to \chi^2(2).$$

## 2.3 OLS Estimation of Autoregression with Time Trend

Now consider a general autoregressive process around a deterministic time trend.

$$y_t = \alpha + \delta t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + u_t,$$

or in matrix form,

$$y_t = x_t' \beta + u_t$$

where $x_t' = [y_{t-1}, y_{t-2}, \ldots, y_{t-p}, 1, t]$, and $\beta' = [\phi_1, \ldots, \phi_p, \alpha, \delta]$. Sims, Stock and Watson (1990) suggest that we find a matrix $G$ and use it to transform this process to

$$y_t = x_t' G' [G']^{-1} \beta + u_t = \tilde{x}_t' \beta^* + u_t.$$

where $\tilde{x}_t = G x_t = [\tilde{y}_{t-1}, \tilde{y}_{t-2}, \ldots, \tilde{y}_{t-p}, 1, t]'$ and $\beta^* = [G']^{-1}\beta = [\phi_1^*, \phi_2^*, \ldots \phi_p^*, \alpha^*, \delta^*]'$.

The idea is that after the transformation, we could write $y_t$ in terms of zero-mean covariance stationary process ($\tilde{y}_{t-j}$), a constant and a time trend. In doing this, we could isolate components of the OLS coefficient vector with different rates of convergence. In this case, after the transformation, $\hat{\phi}_{1,n}^*, \hat{\phi}_{2,n}^*, \ldots$ will converge at the usual rate of $\sqrt{n}$, while $\hat{\alpha}_n^*, \hat{\delta}_n^*$ will behave asymptotically like $\hat{\alpha}_n$ and $\hat{\delta}_n$ in the simple time trend model. The matrix $G$ is of dimension $(p+2) \times (p+2)$:

$$G' \equiv \begin{bmatrix} 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 \\ -\alpha^* + \delta^* & -\alpha^* + 2\delta^* & \ldots & -\alpha^* + p\delta^* & 1 & 0 \\ -\delta^* & -\delta^* & \ldots & -\delta^* & 0 & 1 \end{bmatrix},$$

$$[G']^{-1} = \begin{bmatrix} 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 \\ \alpha^* - \delta^* & \alpha^* - 2\delta^* & \ldots & \alpha^* - p\delta^* & 1 & 0 \\ \delta^* & \delta^* & \ldots & \delta^* & 0 & 1 \end{bmatrix}.$$

The relation between the OLS coefficient estimates before and after the transformation is: $\hat{\beta}_n^* = [G']^{-1}\hat{\beta}_n$ and $\hat{\beta}_n = G'\hat{\beta}_n^*$. A simple example to understand this transformation is the following model:

$$y_t = \phi y_{t-1} + \alpha + u_t, \tag{6}$$

for which we know that $E(y_t) = \alpha/(1-\phi)$. Now, we can write

$$G' \equiv \begin{bmatrix} 1 & 0 \\ -\alpha^* & 1 \end{bmatrix} \quad [G']^{-1} \equiv \begin{bmatrix} 1 & 0 \\ \alpha^* & 1 \end{bmatrix}.$$

We can then solve $\beta^*$ from

$$\begin{bmatrix} \phi^* \\ \alpha^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \alpha^* & 1 \end{bmatrix} \begin{bmatrix} \phi \\ \alpha \end{bmatrix} \Rightarrow \begin{bmatrix} \phi^* \\ \alpha^* \end{bmatrix} = \begin{bmatrix} \phi \\ \alpha/(1-\phi) \end{bmatrix}$$

Then we can rewrite the process $y_t$ as

$$\begin{aligned} y_t &= \phi^* \tilde{y}_{t-1} + \alpha^* + u_t \\ &= \phi\left(y_{t-1} - \frac{\alpha}{1-\phi}\right) + \frac{\alpha}{1-\phi} + u_t \\ &= \phi y_{t-1} + \alpha_t + u_t. \end{aligned}$$

The advantage of this transformation is that now $\tilde{y}_{t-1}$ is a zero mean process. When a time trend is included in the process, we will see similar fact: $\tilde{y}_t$ is demeaned and detrended.

To derive the asymptotic distribution of $\hat{\beta}_n^*$, define

$$H_n = \begin{bmatrix} \sqrt{n} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \sqrt{n} & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{n} & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & \sqrt{n} & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & n^{3/2} \end{bmatrix},$$

then the OLS estimates

$$\hat{\beta}_n^* - \hat{\beta}^* = \left[\sum_{t=1}^n \tilde{x}_t \tilde{x}_t'\right]^{-1} \left[\sum_{t=1}^n \tilde{x}_t u_t\right],$$

$$H_n(\hat{\beta}_n^* - \hat{\beta}^*) = \left[H_n^{-1}\left(\sum_{t=1}^n \tilde{x}_t \tilde{x}_t'\right) H_n^{-1}\right]^{-1} \left[H_n^{-1}\sum_{t=1}^n \tilde{x}_t u_t\right]. \tag{7}$$

Consider the first term and it can be written in the form

$$H_n^{-1}\left(\sum_{t=1}^n \tilde{x}_t \tilde{x}_t'\right) H_n^{-1} = \begin{bmatrix} A_n & B_n \\ B_n' & C_n \end{bmatrix}.$$

The elements in $A_n$ $(p \times p)$ takes the form of $n^{-1}\sum_{t=1}^n \tilde{y}_{t-i}\tilde{y}_{t-j}$ for $i, j = 1, \dots p$, which converges to $\gamma_y^*(|i-j|)$. We can let $Q_{11}^*$ to denote the limiting matrix of $A_n$: $A_n \to_p Q_{11}^*$. Next, the elements in $B_n$ $(p \times 2)$ takes the form of $n^{-1}\sum_{t=1}^n \tilde{y}_{t-i}$ and $n^{-1}\sum_{t=1}^n (t/n)\tilde{y}_{t-i}$, and we know that all of these elements converges to zero: $B_n \to_p 0$. Finally, the matrix of $C_n$ $(2 \times 2)$ is

$$\begin{bmatrix} 1 & n^{-2}\sum_{t=1}^n t \\ n^{-2}\sum_{t=1}^n t & n^{-3}\sum_{t=1}^n t^2 \end{bmatrix} \to \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix} \equiv Q_{22}^*,$$

which is just the $Q$ matrix in our simple time trend model. Thus we have

$$H_n^{-1}\left(\sum_{t=1}^n \tilde{x}_t \tilde{x}_t'\right) H_n^{-1} \to Q^* \equiv \begin{bmatrix} Q_{11}^* & \mathbf{0} \\ \mathbf{0} & Q_{22}^* \end{bmatrix}.$$

Next consider the second term in (7),

$$H_n^{-1}\sum_{t=1}^n \tilde{x}_t u_t = n^{-1/2} \begin{bmatrix} \sum \tilde{y}_{t-1} u_t \\ \sum \tilde{y}_{t-2} u_t \\ \vdots \\ \sum \tilde{y}_{t-p} u_t \\ \sum u_t \\ \sum (t/n) u_t \end{bmatrix} \equiv n^{-1/2}\sum_{t=1}^n \boldsymbol{\xi}_t.$$

This $\boldsymbol{\xi}_t$ is a martingale difference sequence with variance

$$E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') = \sigma^2 Q_t^* \equiv \sigma^2 \begin{bmatrix} Q_{11}^* & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & t/n \\ \mathbf{0} & t/n & t^2/n^2 \end{bmatrix}$$

and we have

$$n^{-1}\sum_{t=1}^{n} Q_t^* \to Q^*.$$

Then apply CLT

$$H_n^{-1}\sum_{t=1}^{n} \tilde{x}_t u_t \to N(0, \sigma^2 Q^*).$$

Therefore for the OLS estimate $\hat{\beta}_n^*$ we have

$$H_n(\hat{\beta}_n^* - \beta^*) \to N(\mathbf{0}, [Q^*]^{-1}\sigma^2 Q^*[Q^*]^{-1}) = N(\mathbf{0}, \sigma^2[Q^*]^{-1}).$$

Using the block-diagonal representation, we can also write

$$[Q^*]^{-1} = \begin{bmatrix} [Q_{11}^*]^{-1} & \mathbf{0} \\ \mathbf{0} & [Q_{22}^*]^{-1} \end{bmatrix}.$$

Now, given the asymptotic distribution of the estimates $\hat{\beta}_n^*$, what are the results for $\hat{\beta}_n$, the estimates for the coefficients in the original model? We have that $\hat{\beta}_n = G'\hat{\beta}_n^*$, or in matrix form

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \dots \\ \hat{\phi}_p \\ \hat{\alpha} \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \\ -\alpha^* + \delta^* & -\alpha^* + 2\delta^* & \dots & -\alpha^* + p\delta^* & 1 & 0 \\ -\delta^* & -\delta^* & \dots & -\delta^* & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\phi}_1^* \\ \hat{\phi}_2^* \\ \dots \\ \hat{\phi}_p^* \\ \hat{\alpha}^* \\ \hat{\delta}^* \end{bmatrix}.$$

Note that the $\hat{\phi}_j$ is identical to $\hat{\phi}_j^*$, so for $\phi_j$, we have

$$\sqrt{n}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}) \to N(\mathbf{0}, \sigma^2[Q_{11}^*]^{-1}).$$

Next, $\hat{\alpha}_n$ is a linear combination of variables that converge to a Gaussian distribution at rate $\sqrt{n}$, so $\hat{\alpha}_n$ behaves the same way. Let

$$g_\alpha' = \begin{bmatrix} -\alpha^* + \delta^* & -\alpha^* + 2\delta^* & \dots & -\alpha^* + p\delta^* & 1 & 0 \end{bmatrix},$$

then $\hat{\alpha}_n = g_a'\hat{\beta}_n^*$,

$$\sqrt{n}(\hat{\alpha}_n - \alpha) \to N(0, g_\alpha'[Q^*]^{-1}g_\alpha).$$

Next, $\hat{\delta}_n$ is a linear combination of variables converging at different rates:

$$\hat{\delta}_n = g_\delta'\hat{\beta}_n^* + \hat{\delta}_n^*$$

where

$$g_\delta = \left[ \begin{array}{cccccc} -\delta^* & -\delta^* & \dots & -\delta^* & 0 & 0 \end{array} \right].$$

Its asymptotic distribution is governed by the variables with the slowest rate of convergence:

$$\begin{aligned} \sqrt{n}(\hat{\delta}_n - \delta) &= \sqrt{n}(\hat{\delta}^*_n + g'_\delta \hat{\beta}^*_n - \delta^* - g'_\delta \beta^*) \\ &\to_p \sqrt{n}(\delta^* + g'_\delta \hat{\beta}^*_n - \delta^* - g'_\delta \beta^*) \\ &= g'_\delta \sqrt{n}(\hat{\beta}^*_n - \beta^*) \\ &\to_p N(0, \sigma^2 g_\delta [Q^*]^{-1} g_\delta). \end{aligned}$$

So each element of $\hat{\beta}_n$ individually is asymptotically Gaussian and $O_p(n^{-1/2})$. The asymptotic distribution of the full vector $\sqrt{n}(\hat{\beta}_n - \beta)$ is multivariate Gaussian, though with a singular variance-covariance matrix.

For hypothesis testing in this model, please read Hamilton's book.