Lecture 2: Time series regression: Further Issues

Yu Bai

City University of Macau

Stationary Time Series

Stationary Stochastic Process

The stochastic process $\{x_t : t = 1, 2, \cdots\}$ is *stationary* if for every collection of time indices $1 \leq t_1 < t_2 < \cdots < t_m$, the joint distribution of $(x_{t_1}, x_{t_2}, \cdots, x_{t_m})$ is the same as the joint distribution of $(x_{t_1+h}, x_{t_2+h}, \cdots, x_{t_m+h})$ for all integers $h \geq 1$.

Covariance Stationary Process

A stochastic process $\{x_t : t = 1, 2, \cdots\}$ with a finite second moment $\mathbb{E}(x_t^2) < \infty$ is **covariance stationary** if

1. $\mathbb{E}(x_t)$ is constant;
2. $\mathbb{V}(x_t)$ is constant;
3. for any $t, h \geq 1$, $\text{Cov}\,(x_t, x_{t+h})$ depends only on h and not on t.

Stationary Time Series

- A stationary process with finite second moments is covariance stationary.
- The converse is not true.

Example

Let $Z_t \overset{i.i.d.}{\sim} N(0,1)$ and define

$$X_t = \begin{cases} Z_t & \text{if t is even} \\ (Z_{t-1}^2 - 1)/\sqrt{2}, & \text{if t is odd,} \end{cases}$$

It is straightforward to verify that $(x_t)_t$ is covariance stationary. However, $(x_t)_t$ can not be stationary as when t is even it is Normally distributed but it follows a χ^2 distribution when t is odd.

- However, if $(x_t)_t$ is a covariance stationary Gaussian process then $(x_t)_t$ is stationary.

Stationary Time Series

- Sometimes, to emphasize that stationarity is a stronger requirement than covariance stationarity, the former is referred to as strict stationarity.
- Because strict stationarity simplifies the statements of some of our subsequent assumptions, "stationarity" for us will always mean the strict form.

Exercise

Suppose that $\{y_t, t = 1, 2, \cdots\}$ is generated by $y_t = \delta_0 + \delta_1 t + e_t$, where e_t is an *i.i.d.* sequence with mean zero and variance σ_e^2.

1. Is $(y_t)_t$ covariance stationary?
2. Is $(y_t - \mathbb{E}(y_t))_t$ covariance stationary?

Weakly Dependent Time Series

- The exact definition requires some technical notions on characterising dependence, but we do not need that.
- For our purposes, an intuitive notion of the meaning of weak dependence is sufficient.
- We say that a time series $\{x_t : t = 1, 2, \cdots\}$ is weakly dependent if $\text{Corr}(x_t, x_{t+h}) \neq 0$ for a fixed h but $\text{Corr}(x_t, x_{t+h}) \to 0$ as $h \to \infty$.
- This asymptotically independence is enough for WLLN and CLT to hold in the time series context.

Example: MA(1) process

$$x_t = e_t + \alpha_1 e_{t-1}, \quad t = 1, 2, \cdots,$$

where

- $(e_t)_t$ is an *i.i.d.* sequence with zero mean and variance σ_e^2.

Let us calculate some quantities:

- $\mathbb{V}(x_t) = (1 + \alpha_1^2)\sigma_e^2$
- $\text{Cov}(x_t, x_{t+1}) = \alpha_1 \mathbb{V}(e_t) = \alpha_1 \sigma_e^2$
- $\text{Cov}(x_t, x_{t+h}) = 0$, for any $h > 1$
- $\text{Corr}(x_t, x_{t+1}) = \frac{\alpha_1}{1+\alpha_1^2}$, which is peaked when $\alpha_1 = 1$

Example: AR(1) process

$$y_t = \rho_1 y_{t-1} + e_t, \ \ t = 1, 2, \cdots ,$$

where

- $(e_t)_t$ is an *i.i.d.* sequence with zero mean and variance σ_e^2;
- y_0 satisfies $\mathbb{E}(y_0) = 0$.

The crucial assumption for weak dependence of an AR(1) process is the stability condition $|\rho_1| < 1$.

- $\mathbb{E}(y_t) = \mathbb{E}(y_{t-1}) \Rightarrow \mathbb{E}(y_t) = 0$
- $\mathbb{V}(y_t) = \rho_1^2 \mathbb{V}(y_{t-1}) + \sigma_e^2 \Rightarrow \sigma_y^2 = \sigma_e^2 / \left(1 - \rho_1^2\right)$
- $\text{Cov}(y_t, y_{t+h}) = \rho_1^h \sigma_y^2$
- $\text{Corr}(y_t, y_{t+h}) = \rho_1^h$

- By repeated substitution, we have

$$\begin{aligned} y_{t+h} &= \rho_1 y_{t+h-1} + e_{t+h} = \rho_1 \left(\rho_1 y_{t+h-2} + e_{t+h-1}\right) + e_{t+h} \\ &= \rho_1^2 y_{t+h-2} + \rho_1 e_{t+h-1} + e_{t+h} = \cdots \\ &= \rho_1^h y_t + \sum_{\ell=1}^{h} \rho_1^{h-\ell} e_{t+\ell} \end{aligned}$$

- So that

$$\begin{aligned} \mathrm{Cov}(y_t, y_{t+h}) = \mathbb{E}(y_t y_{t+h}) &= \rho_1^h \mathbb{E}(y_t^2) + \sum_{\ell=1}^{h} \rho_1^{h-\ell} \mathbb{E}\left(y_t e_{t+\ell}\right) \\ &= \rho_1^h \mathbb{E}(y_t^2) = \rho_1^h \sigma_y^2. \end{aligned}$$

- Something about trend-stationary process and weakly dependence...

Consistency and asymptotic Normality

Definition

Let W_n be an estimator of θ based on a sample $Y_1, Y_2, \cdots, Y_n$ of size n. Then, W_n is a **consistent estimator** of θ if for every $\varepsilon > 0$,

$$\mathbb{P}\left(|W_n - \theta| > \varepsilon\right) \to 0, \tag{1}$$

as $n \to \infty$.

Definition

Let $(Z_n)_n$ be a sequence of random variables, such that for all numbers z,

$$\mathbb{P}\left(Z_n \leq z\right) \longrightarrow \Phi(z), \tag{2}$$

as $n \to \infty$, where $\Phi(z)$ is the standard normal cumulative distribution function. Then, Z_n is said to have an *asymptotic standard normal distribution*.

Consistency of OLS

- Since

$$\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = \left(\sum_{t=1}^{n} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \left(\sum_{t=1}^{n} \mathbf{x}_t u_t\right),$$

consistency requires

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t \mathbf{x}_t' \xrightarrow{p} \mathbb{E}\left(\mathbf{x}_1 \mathbf{x}_1'\right), \quad (3)$$

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t u_t \xrightarrow{p} \mathbb{E}\left(\mathbf{x}_1 u_1\right) = 0, \quad (4)$$

where $\mathbb{E}\left(\mathbf{x}_1 \mathbf{x}_1'\right)$ has to be *p.d*.

- We will not go through the proofs of (3) and (4), but rather state the assumptions which guarantee them.

Assumption TS.1' Linearity and Weak Dependence

We assume the model is exactly as in Assumption TS.1, but now we add the assumption that $\{(\mathbf{x}_t, y_t), t = 1, 2, \cdots\}$ is stationary and weakly dependent. In particular, the LLN and the CLT can be applied to sample averages.

Assumption TS.2' No Perfect Collinearity

Same as Assumption TS.2.

Assumption TS.3' Zero Conditional Mean

The explanatory variables $\mathbf{x}_t$ are contemporaneously exogenous: $\mathbb{E}(u_t|\mathbf{x}_t) = 0$.

Consistency of OLS

Theorem (Consistency of OLS)

Under Assumption TS.1', Assumption TS.2', and Assumption TS.3', the OLS estimators are consistent: $\operatorname{plim}_{n\to\infty} \hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}$.

Going deeper...

Something on predictive regressions..

Example (Static model)

Consider the model

$$y_t = \alpha + \mathbf{z}_t' \boldsymbol{\beta} + u_t.$$

Under weak dependence, the condition sufficient for consistency of OLS is

$$\mathbb{E}\left(u_t | \mathbf{z}_t\right) = 0.$$

This rules out omitted variable bias and misspecification of functional forms. However, u_{t-1} and $\mathbf{z}_t$ can be correlated (feedback).

Example (Finite distributed lag model)

Consider the model

$$y_t = \alpha + \sum_{\ell=0}^{p} \delta_\ell z_{t-\ell} + u_t.$$

A very natural assumption is that the expected value of u_t, given current and all past values of z, is zero:

$$\mathbb{E}\left(u_t | z_t, z_{t-1}, z_{t-2}, \cdots\right) = 0.$$

This means that, once $z_t, z_{t-1}, \cdots, z_{t-p}$ no further lags of z affect $\mathbb{E}\left(u_t | z_t, z_{t-1}, z_{t-2}, \cdots\right)$. If this were not true, we would put further lags into the equation. As in the previous example, Assumption TS.3' does not rule out feedback from y to future values of z.

Example (AR(1) model)

Consider the model

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t,$$

where the error u_t has a zero expected value, given all past values of y:

$$\mathbb{E}\left(u_t | y_{t-1}, y_{t-2}, \cdots\right) = 0.$$

This implies that

$$\mathbb{E}\left(y_t | y_{t-1}, y_{t-2}, \cdots\right) = \beta_0 + \beta_1 y_{t-1}.$$

Thus, once y lagged one period has been controlled for, no further lags of y affect the expected value of y_t.

The Assumption TS.3' is clearly satisfied. What about Assumption TS.3? (left as an exercise)

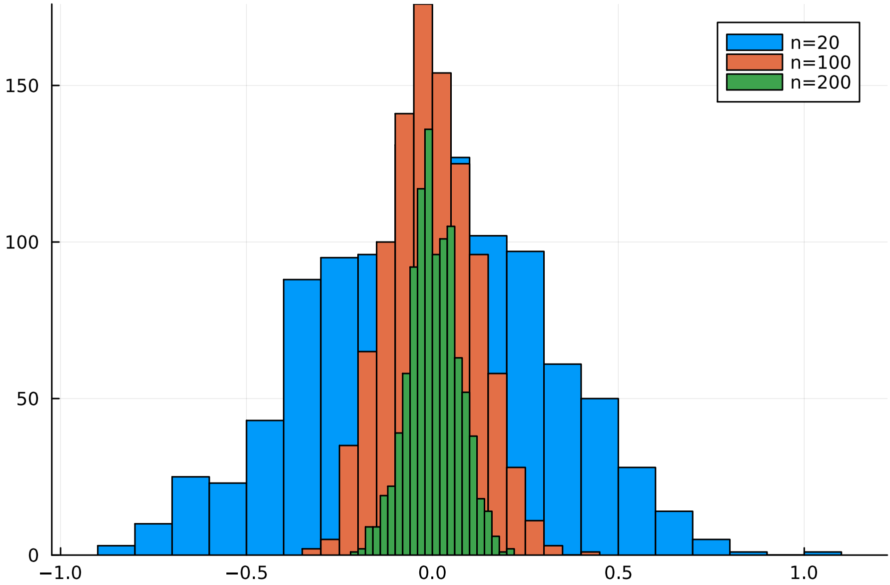
Some simulation exercises

- We generate data according to

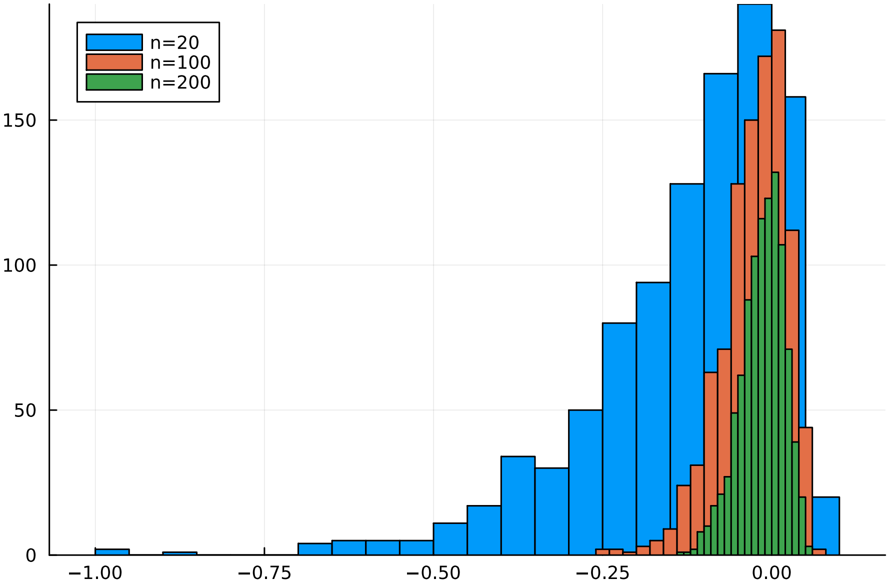
$$y_t = \alpha + \beta X_t + u_t, \quad t = 1, 2, \cdots, n,$$

  where $u_t \sim \mathcal{N}(0, 1)$.
- Two experiments:
  1. (i) $\alpha = 1$, $\beta = 2$, $X_t \sim \mathcal{N}(0, 1)$;
  2. (ii) $\alpha = 0$, $\beta = 0.9$, $X_t = y_{t-1}$.
- Calculate the empirical distribution of $\hat{\beta}_n - \beta$ based on 1000 replications.

(i)



(ii)



Asymptotic Normality of OLS

- We first state two additional assumptions:

Assumption TS.4' Homoskedasticity

The errors are contemporaneously **homoskedastic**, that is, $\mathbb{V}\left(u_t|\mathbf{x}_t\right) = \sigma^2$.

Assumption TS.5' No Serial Correlation

For all $t \neq s$, $\text{cov}\left(u_t u_s|\mathbf{x}_t, \mathbf{x}_s\right) = 0$.

Asymptotic Normality of OLS

- Then, we have

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) = \left(\frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{x}_t u_t\right)$$
$$\xrightarrow{d} \mathcal{N}\left(0, \sigma^2\left(\mathbb{E}(\mathbf{x}_1\mathbf{x}_1')\right)^{-1}\right).$$

- We now summarize the results in the following theorem.

Theorem (Asymptotic Normality of OLS)

Under Assumption TS.1' through Assumption TS.5', the OLS estimators are asymptotically normally distributed. Further, the usual OLS standard errors, t-statistics, F-statistics, and LM statistics are asymptotically valid.

Some simulation exercise (Cont.)

- Consider Case (ii):

$$y_t = \beta y_{t-1} + u_t,$$

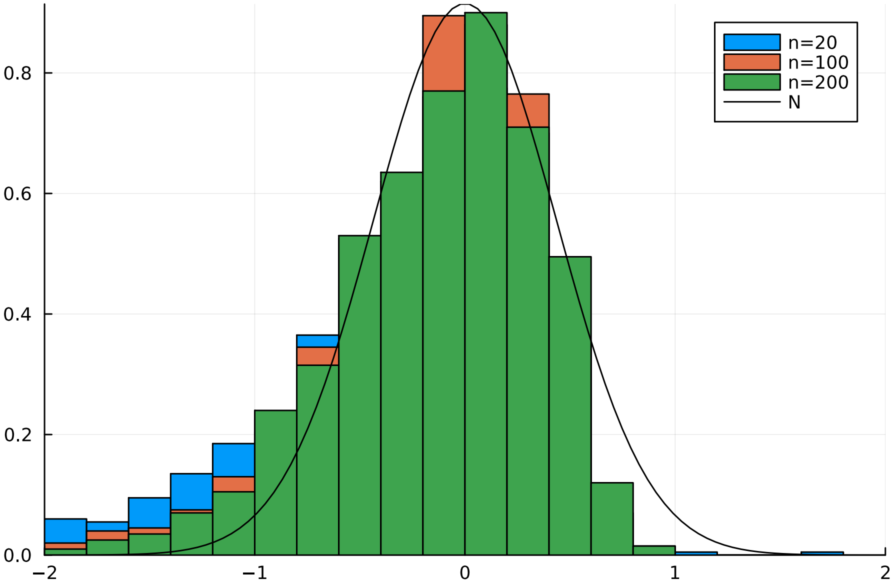
  where $u_t \sim \mathcal{N}(0, \sigma_u^2)$.

- The asymptotic distribution of the OLS estimator $\hat{\beta}_n$ is given by

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, 1 - \beta^2\right).$$

- Let us again compared the empirical distribution of $\sqrt{n}\left(\hat{\beta}_n - \beta\right)$ with the Normal density $\mathcal{N}\left(0, 1 - \beta^2\right)$.
- We set $\sigma_u = 0.2$ and $\beta = 0.9$.

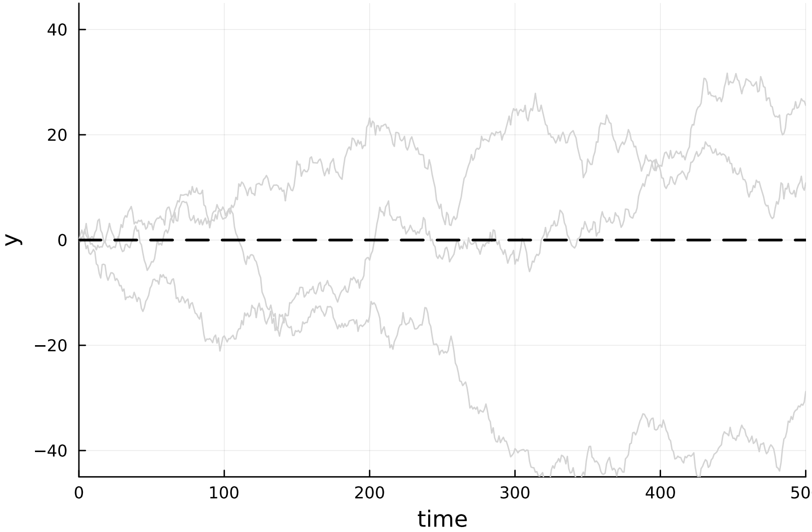
(ii)



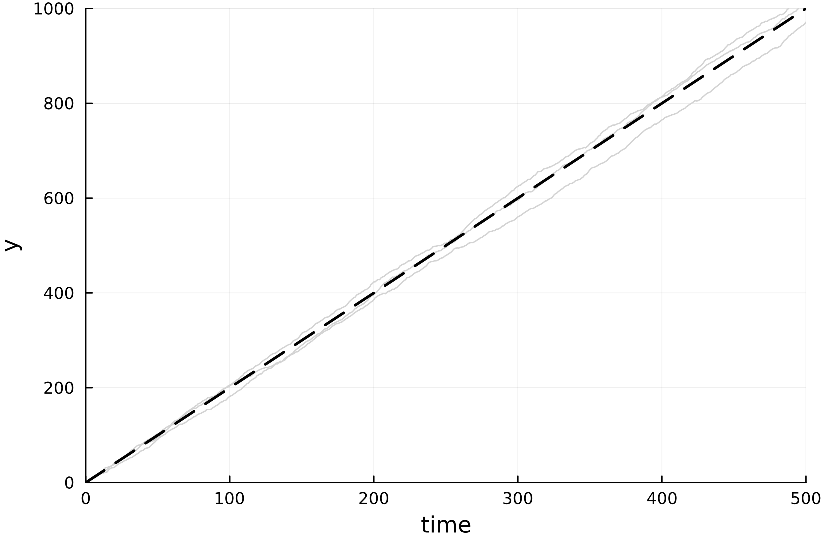
Highly persistent time series

- The process
$$y_t = y_{t-1} + e_t, \ \ t = 1, 2, \cdots,$$
is called a **random walk**.
- By assuming $y_0 = 0$ and $(e_t)_t$ be an $i.i.d.$ sequence with mean zero and variance σ_e^2, we could show that
$$\mathbb{E}(y_t) = \mathbb{E}(y_0) = 0, \ \ \forall t \geq 1$$
$$\mathbb{V}(y_t) = \sigma_e^2 t.$$
- In addition, it can be shown that
$$\text{Corr}\left(y_t, y_{t+h}\right) = \sqrt{t/(t+h)}.$$
- The random walk process is not covariance stationary. Why?
- Can have drift term in the process

Example: pure random walk



Example: random walk with drift



Transformations on Highly Persistent Time Series

- I(1) and I(0) processes
- Transformations:
  - Δy_t
  - growth rate: $\Delta \log(y_t)$, $(y_t - y_{t-1})/y_{t-1}$
- Differencing time series also removes any linear time trend.

Deciding Whether a Time Series Is I(1)

- A **random walk** is a special case of what is known as **unit root process**.
- We will discuss more on testing for unit root in Lecture 4.
- Notice that, testing for I(1) can be done simply by using conventional autocorrelation test. Since $\rho_1 = \text{Corr}\left(y_t, y_{t-1}\right) = 1$ if $\left(y_t\right)_t$ is I(1) but $|\rho_1| < 1$ if I(0).

Dynamically Complete Models and the Absence of Serial Correlation

- Consider the general model

$$y_t = \beta_0 + \mathbf{x}_t'\boldsymbol{\beta} + u_t.$$

- If we assume

$$\mathbb{E}\left(u_t|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, \cdots\right) = 0, \tag{5}$$

  we have

$$\mathbb{E}\left(y_t|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, \cdots\right) = \mathbb{E}\left(y_t|\mathbf{x}_t\right).$$

- If (5) is satisfied, we have a dynamically complete model.
- In other words, whatever is in $\mathbf{x}_t$, enough lags have been included so that further lags of y and the explanatory variables do not matter for explaining y_t.
- Pay attention to the differences between (5) and conditions related to exogeneity

Dynamically Complete Models and the Absence of Serial Correlation

- If we have (5), Assumption TS.5' is satisfied.
- To see this, notice that (*w.l.o.g.* assume $s < t$)

$$\begin{aligned}\mathbb{E}\left(u_t u_s|\mathbf{x}_t, \mathbf{x}_s\right) &= \mathbb{E}\left(\mathbb{E}\left(u_t u_s|\mathbf{x}_t, \mathbf{x}_s, u_s\right)|\mathbf{x}_t, \mathbf{x}_s\right) \\ &= \mathbb{E}\left(u_s \mathbb{E}\left(u_t|\mathbf{x}_t, \mathbf{x}_s, u_s\right)|\mathbf{x}_t, \mathbf{x}_s\right) = 0.\end{aligned}$$

- Dynamically completeness implies no serial correlation.

Going further questions 2

Consider the FDL model:

$$y_t = \alpha + \mathbf{z}_t'\boldsymbol{\beta} + u_t,$$

with MA(1) innovation: $u_t = e_t + \alpha_1 e_{t-1}$. $(e_t)_t$ is an *i.i.d.* sequence with mean zero and variance σ_e^2. Can this model be dynamically complete?

The Homoskedasticity Assumption for Time Series Models

We shall briefly discuss the meaning of the Assumption TS.4' in the following models:

- Static model:

$$y_t = \beta_0 + \beta_1 z_t + u_t$$

- AR(1) model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

- FDL model:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 z_{t-1} + \beta_3 z_{t-2} + u_t$$

Problems

Problem 1

Suppose that the equation

$$y_t = \alpha + \delta t + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t,$$

satisfies the sequential exogeneity assumption.

1. Suppose you difference the equation to obtain

   $$\Delta y_t = \delta + \beta_1 \Delta x_{t1} + \cdots + \beta_k \Delta x_{tk} + \Delta u_t.$$

   Why does applying OLS on the differenced equation not generally result in consistent estimator of the β_j?
2. What assumption on the explanatory variables in the original equation would ensure that OLS on the differences consistently estimates the β_j?
3. Suppose $\delta = 0$. Describe what we need to assume for $\mathbf{x}_t = (x_{t1}, \cdots, x_{tk})$ to be sequentially exogenous. Do you think the assumptions are likely to hold in economic applications?