

## Examen Final (26 de noviembre de 2020)

Dr. Irving Arturo De Lira Salvatierra

Puntos totales 200

- El parcial está dividido en dos secciones, una parte individual y una parte en equipo.
- **La entrega del examen será a más tardar el jueves 26 de noviembre a las 7:00 am.**
- El último día para dudas es el lunes 23 de noviembre.
- Para la sección individual se deben enviar los siguientes archivos (al correo [irvingdelir@gmail.com](mailto:irvingdelir@gmail.com)):
  - El código en R de la Sección 1 (enviar en formato PDF). Usar el siguiente nombre: CU\_Nombre\_S1\_Final. Comentar las respuestas y resultados en el mismo script.
  - La base de datos empleada para poder correr el código.
  - NOTA: correré el código de esta sección, así que deben agregar las librerías que utilizan, comentar las secciones, los resultados, los supuestos y demás información que crean relevante. En caso de recibir un error en el código se considerará como una respuesta incorrecta en absoluto.
- Para la sección en grupos solo deben enviar lo siguiente:
  - PPT
  - El código en R que genera los resultados junto con la base de datos empleada.
  - Se deben formar equipos de 5 a 7 alumnos, que deberán presentar durante la clase del 26 de noviembre los hallazgos.
- **Código de ética: los alumnos tienen prohibido utilizar el mismo código o compartir respuestas, cualquier plagio se penalizará de acuerdo con los lineamientos institucionales del ITAM.**
- El puntaje de cada pregunta se encuentra explícito en el documento

## Sección 1 (130 puntos, cada pregunta vale 10 puntos)

### Programación en R individual

1. Generar una base de datos aleatoria (mantener un seed(12346) con 100 variables y 5,000 observaciones). Generar la base suponiendo que existe una correlación entre las 100 variables y suponiendo una distribución normal multivariada. Se deja a discreción el valor de las medias, varianzas y covarianzas de las 100 variables. Por simplicidad, cada una de estas variables las llamaremos  $x_t^i$ .
2. Generar una variable aleatoria con 5,000 observaciones suponiendo una distribución Cauchy con -2 de parámetro local y 1 como parámetro de escala. Por simplicidad, esta variable la llamaremos  $y_t$ .
3. Calcular la pendiente ( $\hat{\beta}_t$ ) de la regresión cuantílica (nivel de 0.2):  $y_{t+1} = \beta_0 + \beta_i x_t^i + \varepsilon_t^i$ . Es decir, se corre una regresión cuantílica para cada una de las 100 variables ( $x_t^i$ ) tomando como variable dependiente ( $y_{t+1}$ ) la variable calculada en (2), y se guardan los coeficientes estimados. Nota: considerar que las observaciones son datos temporales.
4. Para cada periodo t, correr una regresión  $\bar{x}_t^i = \alpha_0 + f_t \hat{\beta}_t$  y guardar la pendiente  $\hat{f}_t$  estimada. Nota: esta es una regresión cross-sectional para cada t, es decir, se toma como variable dependiente las x's y como regresor las pendientes estimadas.
5. Calcular la regresión cuantílica (nivel de 0.2):  $y_{t+1} = \gamma_0 + \gamma_1 \hat{f}_t + \delta_t$ . Con ello, **hemos realizado un análisis de factores de tres pasos**.
6. Explicar la intuición de  $\hat{f}_t$ .
7. Utilizando ggplot graficar  $y_{t+1}$ , el cuantil 0.2 de  $y_{t+1}$  y  $\hat{y}_{t+1}$  generado en (6).
8. Partir la muestra de 5,000 observaciones en 2: una muestra de 1 a 4,000 y otra de 4,001 a 5,000. Repetir (2)-(5) para la primera muestra. Calcular el error de pronóstico, es decir la suma de los errores al cuadrado (llamaremos este análisis dentro de la muestra).
9. Suponiendo una ventana fija de 4,000, calcular los pronósticos para cada una de las observaciones de la segunda muestra y calcular el error de pronóstico (llamaremos este análisis fuera de la muestra).
10. En un solo gráfico mostrar  $y_{t+1}$ , el pronóstico dentro de la muestra y el pronóstico fuera de la muestra.
11. Calcular una regresión simple de  $y_{t+1}$ , versus todos los  $x_t^i$  para un análisis dentro de la muestra y uno fuera de la muestra.
12. Comparar mediante una prueba si hay evidencia de que el modelo de factores es mejor que una regresión lineal para el análisis dentro y fuera de la muestra.
13. Utilizando la base de datos del FMI (WEO October 2020), utilizar la información anual de México de 1980 a 2019 para pronosticar el crecimiento del PIB de 2020 a 2025.
  - a. Para generar el pronóstico usar la metodología de factores (suponiendo que en lugar de una regresión cuantílica son regresiones lineales) tomando como variable dependiente el crecimiento del PIB de México (Gross domestic product, constant prices) y como regresores: Total investment, Gross national savings, Inflation, average consumer prices, Volume of imports of goods and services, Volume of exports of goods and services, Unemployment rate, General government revenue (% GDP), General government gross debt (%GDP).

- b. Usando únicamente la serie de tiempo de crecimiento del PIB, obtener el mejor modelo ARIMA, generar el pronóstico del PIB de 2020 a 2025.
- c. Comparar los resultados dentro de la muestra de (a) y (b) mediante un test.
- d. Graficar los pronósticos de cada modelo y los reportados en el WEO 2020, ¿Cuál de los modelos es más cercano a las proyecciones del FMI?

## **Sección 2 (70 puntos)**

En equipos de 6 personas preparar una presentación en la cual se conteste lo siguiente:

1. (20 puntos) Evolución del COVID-19 en México
2. (20 puntos) Segmentos por sociodemográficos
3. (20 puntos) Impacto en millones de pesos sobre la economía de México, ¿Qué porcentaje del PIB representa?
4. (10 puntos) Pronóstico de muertes y contagios a nivel agregado de diciembre 2020 a marzo 2021.