**Clase 20/01/2023**

**Temas de Clase**

* Regresión Lineal Simple
* Regresión Lineal Múltiple
* Series de tiempo
  + Modelos ARIMA (Predecir Rentabilidad)
  + Modelos GARCH (Predecir Volatilidad)
* VaR (Value at Risk): Z\_a\* Rentabilidad\* Volatilidad

¿Qué buscar? Industria (Predicciones) o Academia (Investigación).

¿Causalidad entre X y Y? Teoría Económica o Finanzas

* Estimación de los parámetros del modelo.
* Signo de los parámetros.
* El modelo puede entregar resultados no esperado o ilógicos.
  + Consideré una forma funcional no correcta.
  + Omisión de variables relevantes.
  + Gran concentración en cierto valor de la variable X.
  + Valores atípicos que me llevan a resultados no esperados.
* Significancia de la variable (Prueba t)

**SLM**

El modelo de regresión lineal simple es:

Y: Variable dependiente; X: Variable independiente; Beta\_0: Intercepto; Beta\_1: Pendiente. Épsilon: Error.

Métodos de estimación de parámetros: Máxima verosimilitud (Estadística) MCO (Econometría).

El error se considera todo lo que explica la variable independiente que no lo explica el modelo.

**Supuestos del Modelo de Regresión:**

* Linealidad en los parámetros (Los betas de grado 1).
* Variable X es conocida (No estocástica).
* La media de los errores es 0
* Homocedasticidad (Varianza de los errores constante)
* Los errores son incorrelacionados (No autocorrelación).
* Los errores tienen una distribución normal (Media 0 Varianza Sigma^2).

**MCO:**

Busca minimizar la suma de los errores cuadrados (LS).

**Clase 21/01/2023**

Interpretación de los betas: Cuando el valor de X varía en 1 unidad, se espera con un nivel de confianza de (1-alpha)\*100%, en que el verdadero valor promedio de Y va a aumentar (o disminuir, depende el signo) entre LI (Límite Inferior) y LS (Límite Superior).

Interpretación del Ejemplo: Por cada PP el riesgo anual de un portafolio se espera, con una confianza del 96%, que la rentabilidad anual se incremente entre 0.37% y 0.53%.

Por cada 100 dólares que se inviertan en gastos de publicidad, se espera que con una confianza del 95%, que las ventas semanales se incrementen entre 919 y 1375 dólares.

Para determinar si una variable es significativa se realiza una prueba de hipótesis donde H0: B=0 y se rechaza si el valor-p es menor al nivel de significancia.

Grados de libertad: # Número de Observaciones (n) - # Número de Parámetros (k).

Rechazar H0 indica que la variable X\_i tiene efecto lineal sobre los cambios de Y. Si no se rechaza H0 indica que la variable no es significativa en el modelo.

R-squared (Coeficiente de determinación) (R^2) (Bondad de Ajuste): El porcentaje de la variabilidad de Y que es explicado por la variación de X.

Interpretación R^2 del Ejemplo: La variabilidad anual de un portafolio es explicada en un 70.3% por la variabilidad del riesgo.

**Modelo de Índice Único:**

Si beta > 1 el activo es más riesgoso que el mercado.

En dicho modelo:

R^2: Riesgo de Mercado; 1-R^2: Riesgo Específico.

Coeficiente de Correlación (Gamma):

Disclaimer: NO CALCULAR LA CORRELACION A DATOS CON TENDENCIA.

**Clase 27/01/2023**

Supuestos:

1. Media de los errores igual a 0: Si la constante no es significativa, de todos modos, se recomienda no eliminarla del modelo para no afectar la media de los residuos.
2. Homocedasticidad: Se realiza el test de White con homocedasticidad como hipótesis nula y heterocedasticidad como hipótesis alternativa que establece que si el valor-p es mayor o igual es mayor que alfa NO se rechaza la hipótesis nula (homocedasticidad). Si se llega a Rechazar la hipótesis nula los estimadores que se calculen pueden ser NO eficientes (varianza alta) y, por ende, que el error típico sea mas grande y la longitud de los intervalos es mayor. Para evitar los valores atípicos se utilizan variables duumies que reemplazan dichos valores atípicos por un 1.
3. Los errores deben ser incorrelacionados: Hacer test de Durbin Watson (DW): Si 1.8<= DW <= 2.2 no existe autocorrelación de primer orden en los residuos, pero podría existir cualquier otro orden de autocorrelación. Si 1.5<=DW<=1.8 o 2.2<=DW<=2.5. NO DICE NI CHIMBA. Si 0<=DW<=1.5 existe autocorrelación positiva de primer orden en los residuos y no se valida el supuesto. Si 2.5<=DW<=4 existe correlación negativa de primer orden y no se valida el supuesta.

También puede realizar la prueba de autocorrelación serial donde H\_0 es que los errores son incorrelacionados. No validar este supuesto implica que el R^2 sea sobrestimado, los estimadores B0 y B1 no son eficientes y las pruebas T y F no son confiables. Para solucionar dicho problema se puede considerar una estructura autorregresiva de primer orden en los residuos del modelo.

1. La distribución de los errores es normal: Se realiza el test de Jarque-Bera que cuando no se rechaza H\_0 se valida el supuesto. De no cumplirse, pierden validez las pruebas T y F. Se soluciona transformando las variables.

**Selección del mejor modelo:**

Si los modelos tienen diferente variable dependiente se calcula en cada modelo la correlación que existe entre los datos reales de Y y los pronósticos. SE SELECCIÓNA EL MODELO QUE TENGA MAYOR CORRELACIÓN.

Si se elimina la constante (porque no es significativa) no se puede garantizar que la media de los errores sea 0. Entonces no se recomienda eliminarla del modelo.

**Clase 03/02/2023**

DISCLAIMER SOBRE EL USO DE LOS RESIDUOS: SE USAN LOS RESIDUOS ESTANDARIZADOS PARA DETECTAR LA PRESENCIA DE OUTLIERS. DE RESTO SE USAN LOS RESIDUOS NORMALES.

* Residuos: error
* Residuos estandarizados: errores/desviación de los errores. De esta manera, los residuos estandarizados toman valores entre -3 y 3.

Para realizar los intervalos de a partir de un valor inicial de X NO debo incluir la dummies. A menos que el X sea el mismo que el outlier que utilizamos para crear la dummy.

**Regresión Múltiple**

Supuestos:

* Las variables X1, X2, X3,…, Xk deben ser independientes entre sí. Si falla este supuesto, el modelo tendría problemas de multicolinealidad (perfecta o aproximada).
  + Hay señales que me ayudan a detectar una posible multicolinealidad aproximada, tales como:
    - R^2 alto, variables individuales no son significativas y la prueba de significancia conjunta es significativa.
    - Si el modelo tiene solo 2 variables explicativas, calcule la correlación entre ellas. Y, si la correlación, en valor absoluto, es mayor o igual a 0.8 entonces existe multicolinealidad aproximada. Disclaimer: si el modelo tiene más de 2 variables explicativas, también se puede calcular la matriz de correlación. Correlaciones bajas NOOOOO descartan multicolinealidad. Se debe calcular el determinante de la matriz de correlación, si este es 1 o muy cercano a 1, las variables X1, X2,…, Xk, son ortogonales (independientes).
    - Calcular el factor de la varianza (VIF): para hacer esto se deben calcular las regresiones de cada variable independiente (como si fuera dependiente) frente a las otras independientes. Si el VIF\_i es mayor o igual a 10, se dice que la variable X\_i está altamente correlacionada con el resto.
  + Los programas no dejan correr la regresión si hay una multicolinealidad perfecta.

Las consecuencias de tener un modelo con multicolinealidad son: Los estimadores no son eficientes, la varianza del error es más grande, la varianza de los estimadores es más grande, se pueden considerar variables no significativas cuando en realidad lo son, los estimadores son muy sensibles a pequeños cambios en la muestra.

Algunas soluciones pueden ser: eliminar las variables colineales si no son estadísticamente significativas; usar análisis de componentes principales (ACP) ya que permiten pasar de un espacio LD a un espacio LI.

Si el límite inferior es negativo y el superior es positivo, la variable NO es significativa en el modelo.

**Clase 04/02/2023**

Prueba de subconjuntos: Se pueden coger las variables que uno quiera pero el profe recomienda coger las que resulten NO significativas para determinar si hay al menos 1 que sea importante en el modelo y no eliminarla arbitrariamente.

* Rechazar H0: Al menos 1 de las variables especificadas es significativa.
* No rechazar: El modelo reducido (el modelo sin las variables no significativas) es estadísticamente significativo.

Se hace una prueba F para subconjuntos especificada como:

Donde: SSE\_R: Suma de cuadrados del error del modelo reducido; SSE\_C: Suma de cuadrados del error del modelo completo; L: es el número de parámetros que hay en H\_0; S\_C^2 es la varianza del modelo completo.

La prueba se rechaza si:

Donde a: Alpha; L: el número de parámetros que hay en H\_0; (n-k-1) \_c: los grados de libertad.

Luego de determinar si hay, o no, variables significativas calculadas como no significativas me dispongo a eliminar variables del modelo empezando por la que tenga mayor valor p.

R^2 ajustado: castiga la inclusión de variables basura en el modelo, al contrario que el R^2 normal que incrementa sin importar la calidad de las variables incluidas en el modelo.

Para realizar el intervalo del pronóstico:

**Clase 10/02/2023**

**Series de Tiempo**

Para clasificar o identificar una serie de tiempo se traza el correlogreama y el grafico de la serie de tiempo.

Correlograma: Es el gráfico de los valores de autocorrelación de una serie de tiempo.

Para calcular el correlograma y el correlograma parcial debe entrar por Risk Simulator, herramientas de análisis, Análisis de Estadísticas (opción 24), y deja seleccionado únicamente el argumento de autocorrelación. Excel abrirá un archivo temporal donde mostrará el grafico de la ACF y PACF. Se puede determinar si una serie es estacionaria si hay ‘pocos’ valores que se salgan de las bandas del correlograma.

Componentes de una serie de tiempo:

1. Si las observaciones fluctúan alrededor de un eje horizontal, se dice que la serie de tiempo es estacionaria en media. La mayoría de los modelos econométricos son diseñados para este tipo de comportamiento (ausencia de tendencia).

Algunas técnicas muy útiles para este tipo de datos pueden ser: Media Móvil, Suavizamiento Exponencial, Modelos AR, Modelos MA, o Modelos ARMA.

1. Si las observaciones presentan un comportamiento creciente o decreciente en el largo plazo, la serie de tiempo tiene tendencia. El correlograma de una serie con tendencia muestra una alta correlación entre los valores sucesivos de la serie.

La tendencia estocástica es la que muestra fuertes fluctuaciones. También hay tendencia determinística que se modela con respecto al tiempo. Algunas técnicas de pronóstico para esta clase de datos son: Doble Media Movil, Doble Suavizamiento Exponencial, Líneas de Tendencia (se usa más para la tendencia determinística), Modelos ARIMA (Tendencia Estocástica).

1. Si las observaciones se repiten regularmente en ciertas épocas del año, se dice que la serie de tiempo tiene un componente estacional. El correlograma de una serie estacional es claro porque en el rezago estacional (número de veces que se observa la serie por año) y/o múltiplos de dicho rezago es significativa.

Algunas técnicas de pronóstico pueden ser: (Para series Estacionales Estacionarias) Regresión Múltiple con Variables Dummy, Aditivo Estacional, Multiplicativo Estacional, SARMA; (Para series Estacionales con Tendencia) Holt Winter Aditivo, Holt Winter Multiplicativo, Tendencia Lineal con Variables Dummy, SARIMA.

1. Si las observaciones no tienen un determinado patrón se dice que la serie de tiempo es aleatoria o Ruido Blanco. El ACF de un ruido blanco no es significativo en ningún rezago; Los datos son incorrelacionados, Los datos no tienen memoria, No existe dependencia lineal entre los datos; Son impredecibles a partir de su propio pasado; Podrían ser predecibles a partir de otras variables. Si todos los valores-p del ACF/PACF son mayores o iguales que Alpha indica que los datos que se están analizando son ruido blanco; si al menos 1 valor-p es menor que Alpha, los datos NO son ruido blanco.

VaR (Diferente de VAR): Value at Risk es la máxima perdida esperada en una inversión en un horizonte de tiempo dado con cierto nivel de confianza.

Aproximaciones a VaR:

* Suponiendo normalidad:
* T-Student:
* Simulación Histórica Filtrada:

La volatilidad se puede estimar con modelos EWMA o GARCH(1,1).

El retorno se puede estimar con modelos AR, MA o ARMA.

* Modelo EWMA (Promedio Móvil con Ponderación Exponencial):

Muy utilizado por JP, Morgan. Risk Metrics.

Lambda toma valor de 0.94 para datos diarios, 0.97 o 0.98 para datos mensuales.

* Semilla: Considere la varianza muestral de los retornos (incorrelacionados).

Supuestos: los retornos deben ser ruido blanco. Si no son ruido blanco se debe encontrar una estructura ARMA para los retornos con el fin de eliminar la dependencia lineal que existen entre ellos.

* Modelo GARCH (1,1) (Engle y Bullerslev).
* Modelo AR (p) (Autorregresivo de Orden P).
* Modelo MA (q) (Media Móvil de Orden q).

**Pasos para resolver el Excel Sheet: USDCOP.**

1. Hallar los retornos de la TRM.
2. Ver si los retornos son ruido blanco.
   1. Los retornos NO son ruido blanco porque el valor-p de su correlograma son menores que Alpha.
   2. Son estacionarios en media ya que pocos valores del correlograma salen de las bandas.
3. Me fijo en que rezago se encuentra el primer valor-p menor que Alpha, observo que está en el rezago 1 entonces inicio estimando un AR(1). Si el primer valor-p menor que Alpha es diferente de 1 me fijo en los rezagos anteriores y selecciono el valor de autocorrelación más alto en valor absoluto y lo incluyo en el modelo.
4. Validación del modelo: Los errores del modelo deben ser ruido blanco. Si no son ruido blanco indica que al modelo se le debe adicionar rezagos de la variable dependiente y/u otras variables.
5. Como no son ruido blanco incluyo el rezago con mayor valor de autocorrelación antes del primer valor-p menor que Alpha. Iterar este paso hasta tener residuos ruido blanco.
6. Guardar los residuos del modelo seleccionado y aplicar las fórmulas de las diferentes metodologías para estimar el VaR.

**Para el modelo GARCH**

Primero se valida el modelo GARCH (1,1). Si los cuadrados de los errores NO son ruido blanco, estimo un orden diferente de GARCH.