

# Modelo de predicción de demanda de gas natural por subsector industrial

*Construcción y análisis de Modelos*

Elaborado por Gabriela Olivera y Fernanda Mora  
Enero 13, 2017

SENER  
SECRETARÍA DE ENERGÍA



# Contenido

- Objetivo
- Consideraciones para el modelo
- Modelos
  - Minería de datos
  - ARIMA
- Siguietes pasos

# Objetivo

# Objetivo

- Construir un **modelo de predicción** de demanda de gas natural para el sector industrial por subramo
- Las **variables sugeridas** en el modelo son:
  - Demanda histórica gas natural (millones de ft cúbicos)
  - Precio gas natural (USD/MMBtu)
  - Precio combustóleo (USD/MMBtu)
  - PIB (miles de millones de pesos)
- El **modelo exitoso** será el que tenga el menor **error de predicción**
- Se usó el **modelo del IMP** como benchmark

# Consideraciones

# Consideraciones

- Los **subsectores industriales** considerados después del análisis y homologación de cuatro fuentes de información son:
  - Industria Alimentaria
  - Fabricación de pulpa, papel y cartón
  - Productos a base de minerales no metálicos
  - Cerveza y malta
  - Metales básicos
  - Productos metálicos, equipo eléctrico y de transporte
  - Productos de minerales no metálicos
  - Química
  - Resto
  - Textil
- Período datos históricos (reales): 2005-2015
- Período modelo de IMP: 2005-2030
- Los **modelos** se realizaron en el **lenguaje R**

# Estructura base de datos de entrada

Año	Categoría unificada (subramo)	Precio industrial de gas natural (PEMEX)	Precio de combustóleo (PEMEX)	PIB	Factores de eficiencia
XXX	XXX	XXX	XXX	XXX	XXX
XXX	XXX	XXX	XXX	XXX	XXX
...	...	...	...	...	...

# Modelos



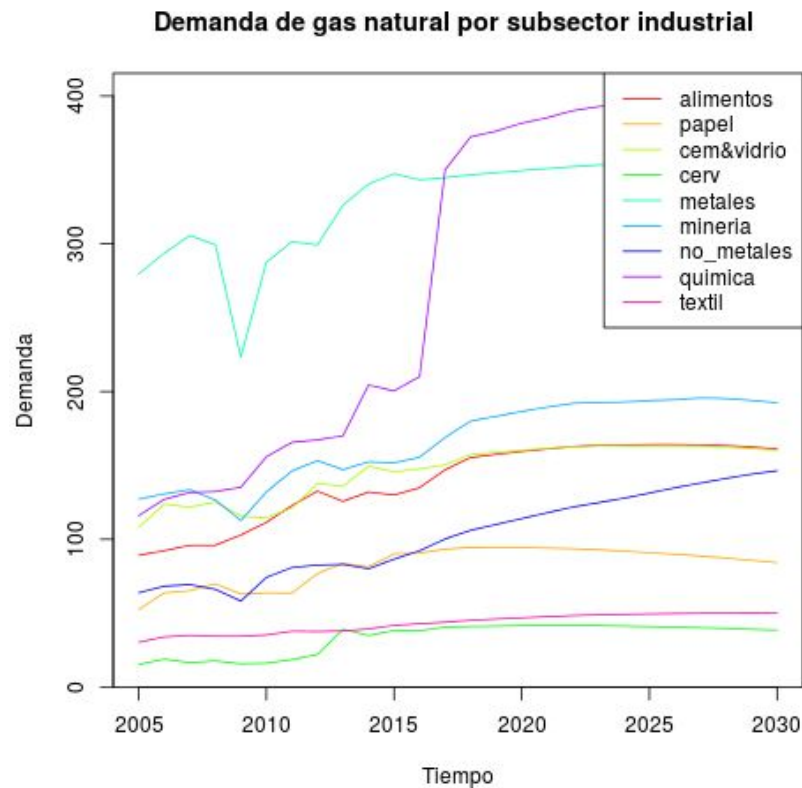
# Ventajas de desarrollar los modelos

- Contar con un modelo con **poder predictivo** que permita generar analíticamente las **prospectivas de demanda**
- Contar un **modelo propio** que no sea una caja negra
- Contar con un modelo **flexible**:
  - Posibilidad de agregar más períodos de tiempo
  - Posibilidad de cambiar granularidad
  - Posibilidad de agregar variables nuevas
  - Posibilidad de replicarlo para cualquier otro tipo de energía
- Contar con **varios modelos** y posibilidad de compararlos para seleccionar el **mejor**

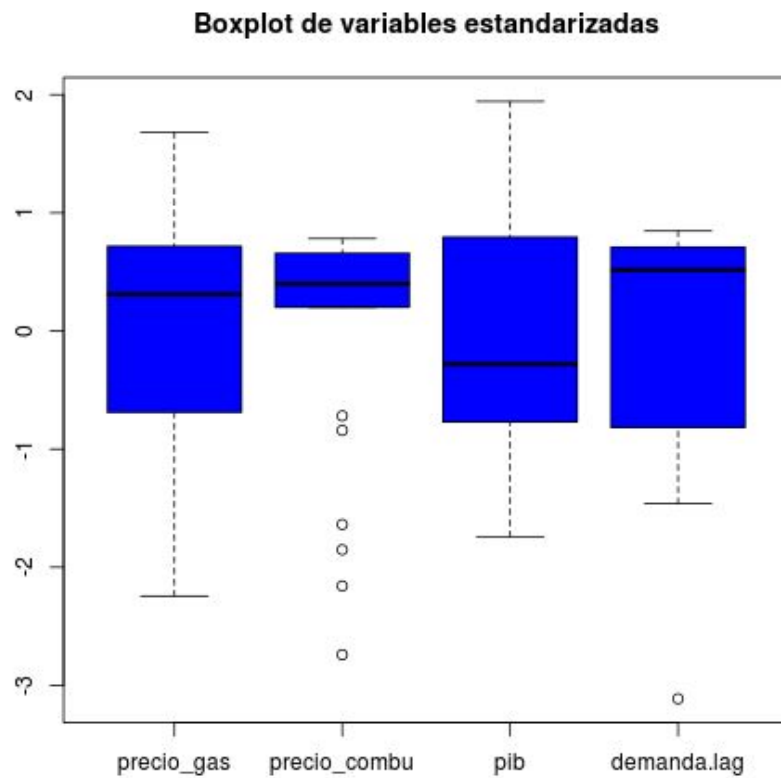
# Planteamiento

- Modelo de **series de tiempo**
- 2 enfoques:
  1. **Estadístico:** Modelos ARIMA (p,d,q), i.e. autoregressive integrated moving average  
**Ventajas:** funciona bien con pocos datos.  
**Desventajas:** paramétrico, supuestos, conocimiento.
  2. **Minería de datos/Machine learning:** regresión lineal múltiple, máquina de soporte, red neuronal.  
**Ventajas:** muchos datos, variedad de algoritmos.  
**Desventajas:** caja negra, mal uso.

# Demanda por sector



# Boxplot



# Minería de datos

# Proceso Modelo Minería de Datos

## Proceso:

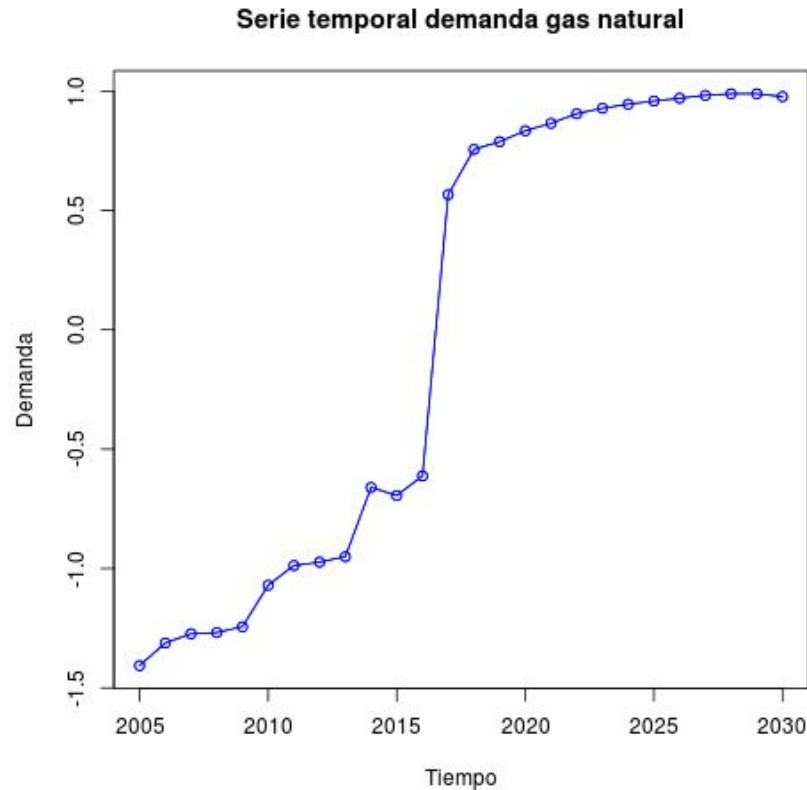
1. Visualizar la serie y estandarizar los datos
2. Análisis descriptivo
3. Reorganizar los datos
4. Construir el modelo de Minería de datos
5. Predecir
6. Evaluar

# Proceso Modelo Minería de Datos

Proceso:

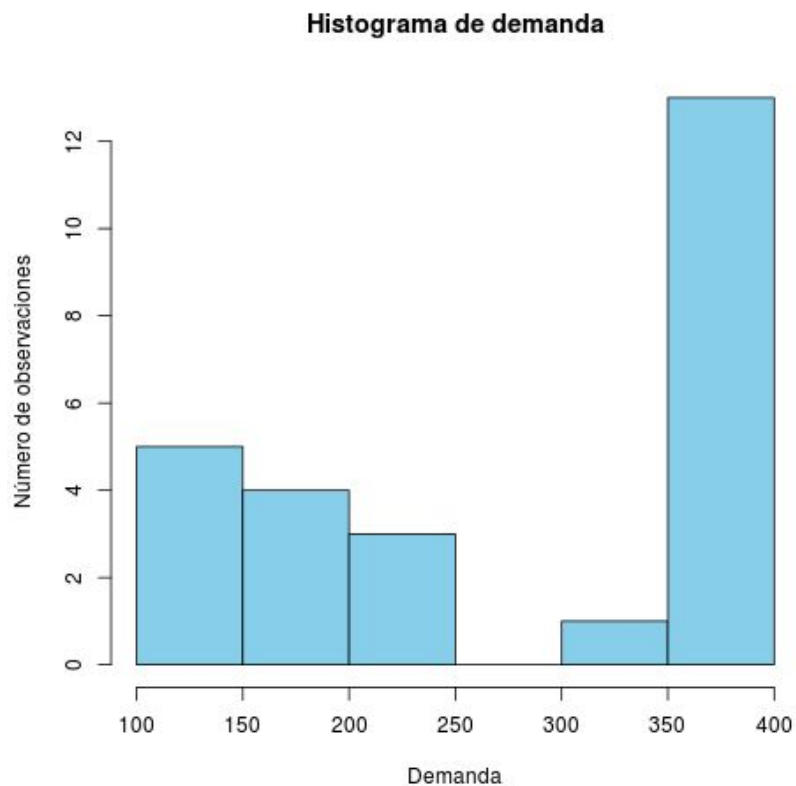
1. Visualizar la serie y estandarizar los datos
2. Análisis descriptivo
3. Reorganizar los datos
4. Construir el modelo de Minería de datos
5. Predecir
6. Evaluar

# Demanda sector Químico (estandarizada)

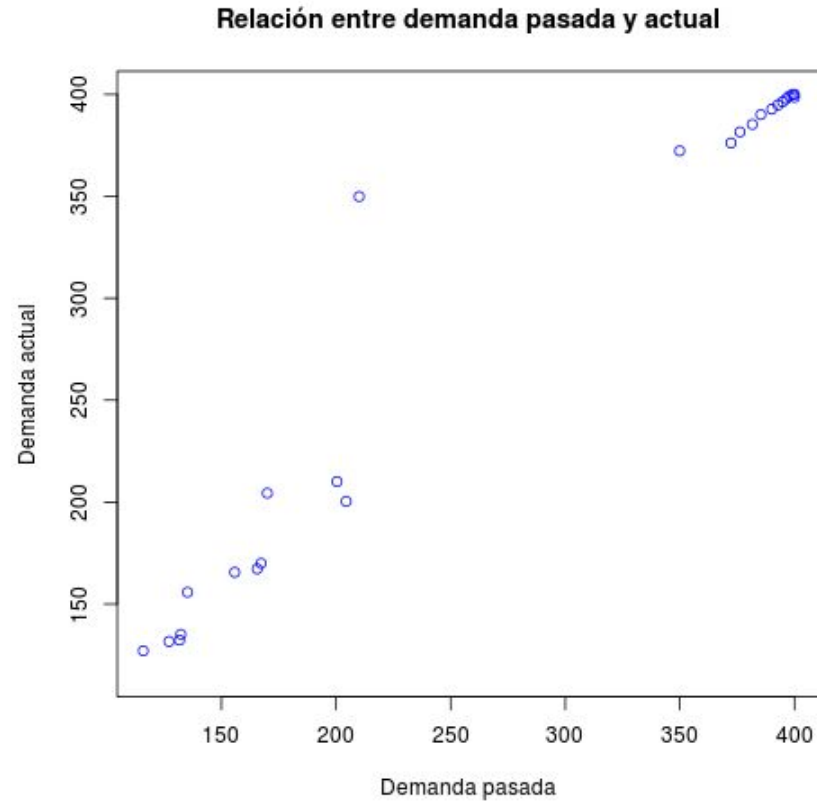




# Histograma sector Químico

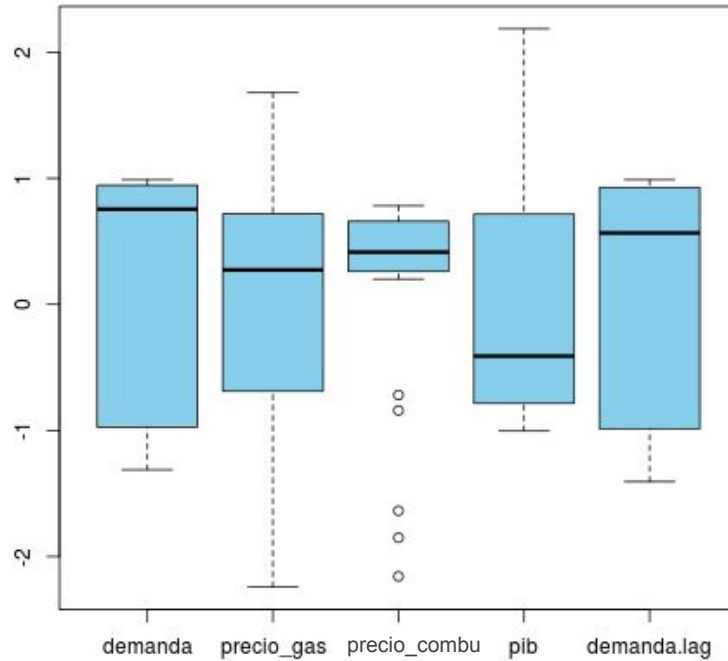


# Demanda sector Químico

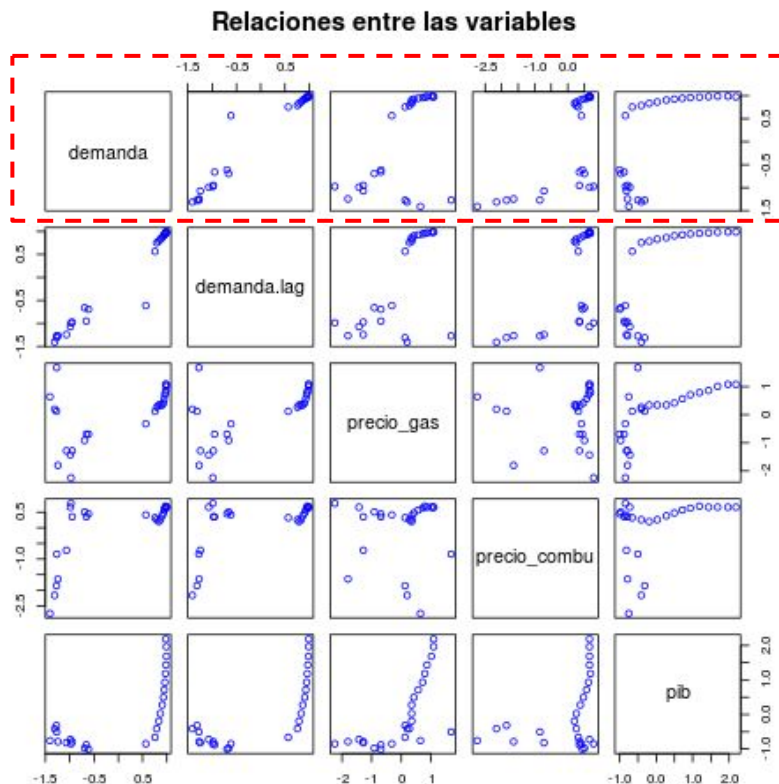


# Boxplot

Boxplot de variables estandarizadas en subsector

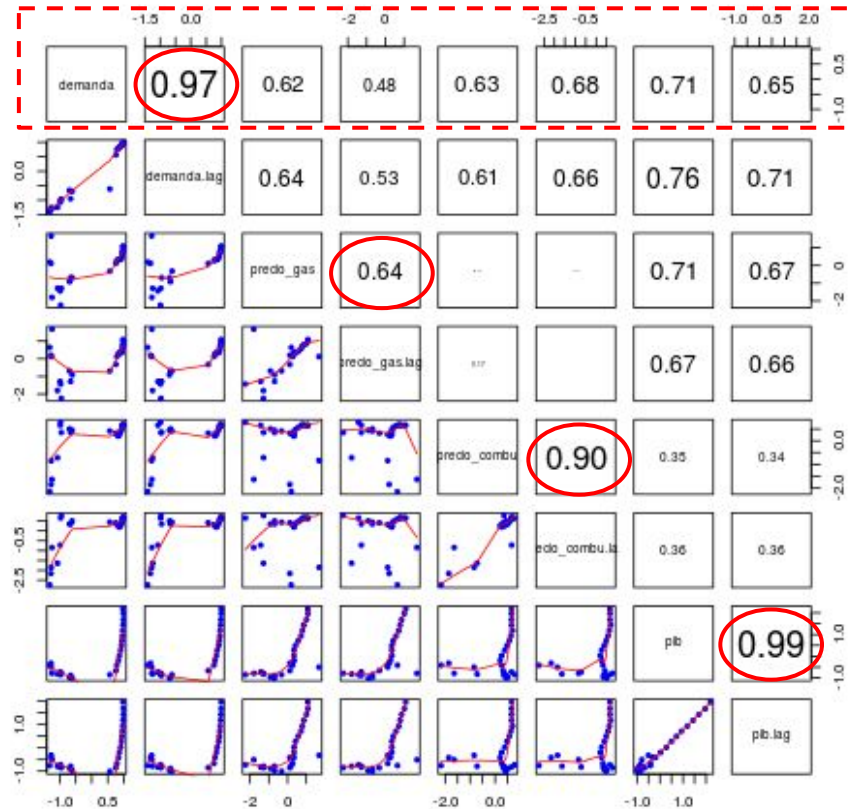


# Scatterplot por par de variables



# Correlaciones por par de variables

Matriz de correlaciones y scatterplots para todas las variables



# Proceso Modelo Minería de Datos

Proceso:

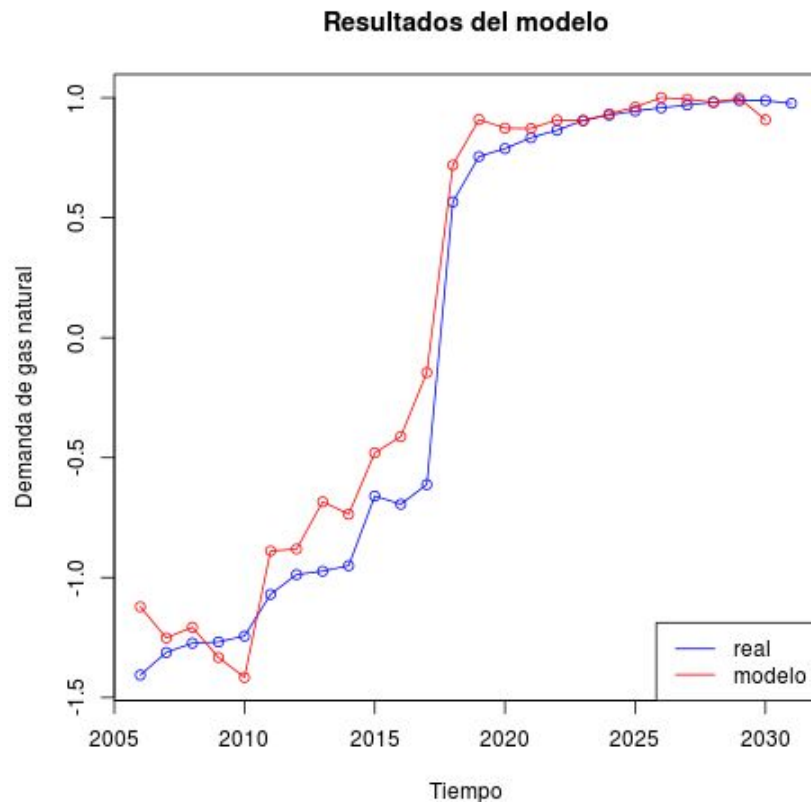
1. Visualizar la serie y estandarizar los datos
2. Análisis descriptivo
3. Pruebas para reorganizar los datos
4. Construir el modelo de Minería de datos
5. Predecir
6. Evaluar

# Proceso Modelo Minería de Datos

Proceso:

1. Visualizar la serie y estandarizar los datos
2. Análisis descriptivo
3. Pruebas para reorganizar los datos
4. Construir el modelo de Minería de datos
5. Predecir
6. Evaluar

# Modelo y variables importantes



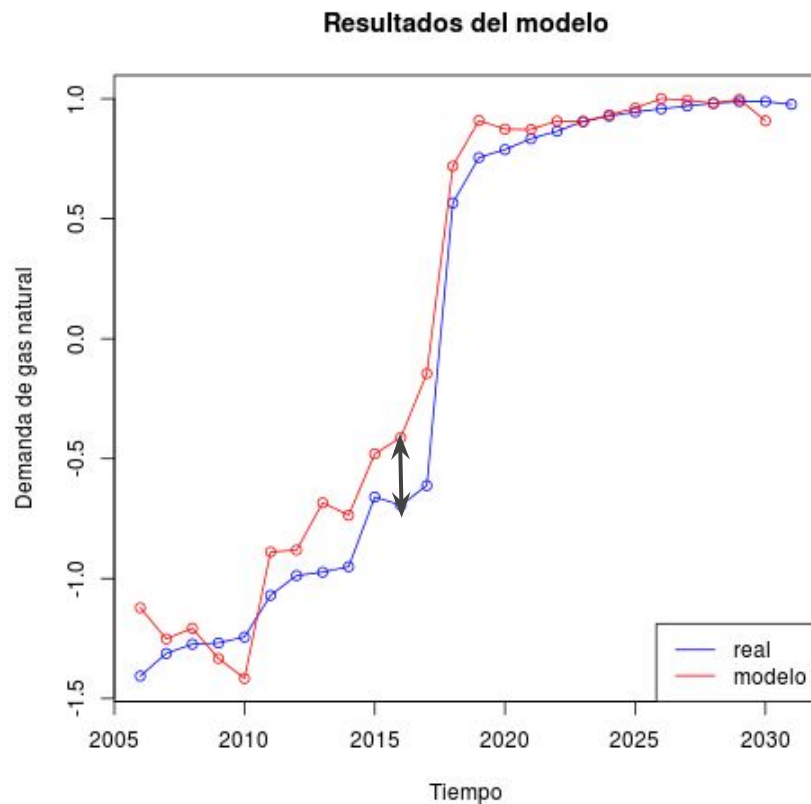
**Modelo:** regresión lineal múltiple

Prueba de hipótesis t sobre los coeficientes:

- Demanda.lag
- Precio gas
- Precio gas.lagged
- Precio combustóleo
- Precio combustóleo lagged
- PIB
- PIB.lagged



# Resultados



Error 2006 - 2015: 0.3

Error 2016 - 2030: 0.58

-----  
Error 2006 - 2030: 0.88

ARIMA

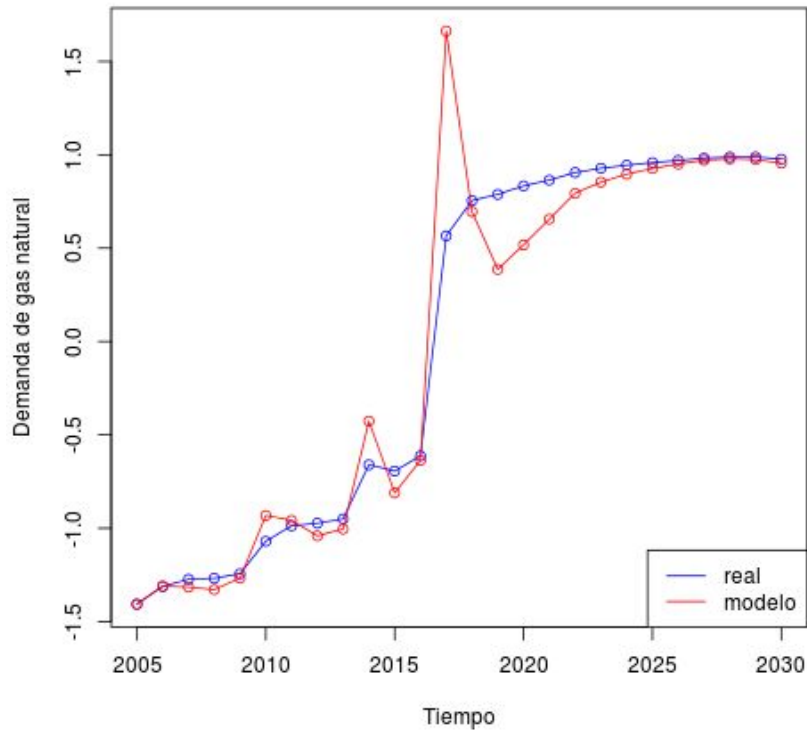
# Proceso Modelo ARIMA

## Proceso:

1. Visualizar la serie y estandarizar los datos
2. Hacerla estacionaria
3. Analizar ACF/PACF
4. Construir el modelo ARIMA
5. Predecir
6. Evaluar

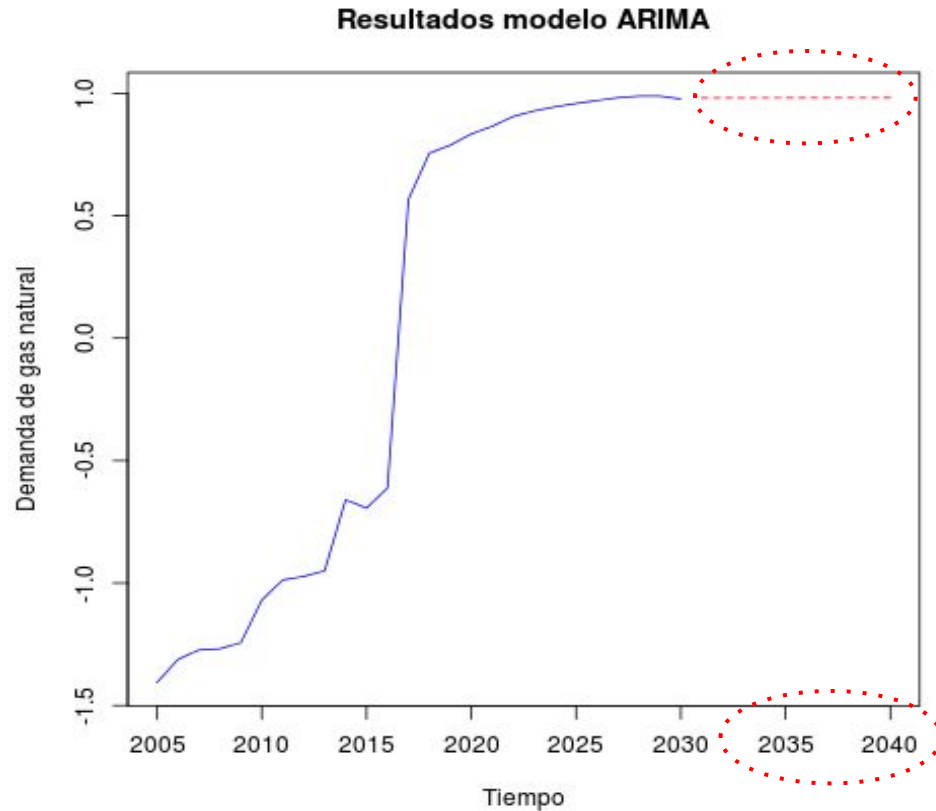
# Resultados

Resultados modelo ARIMA

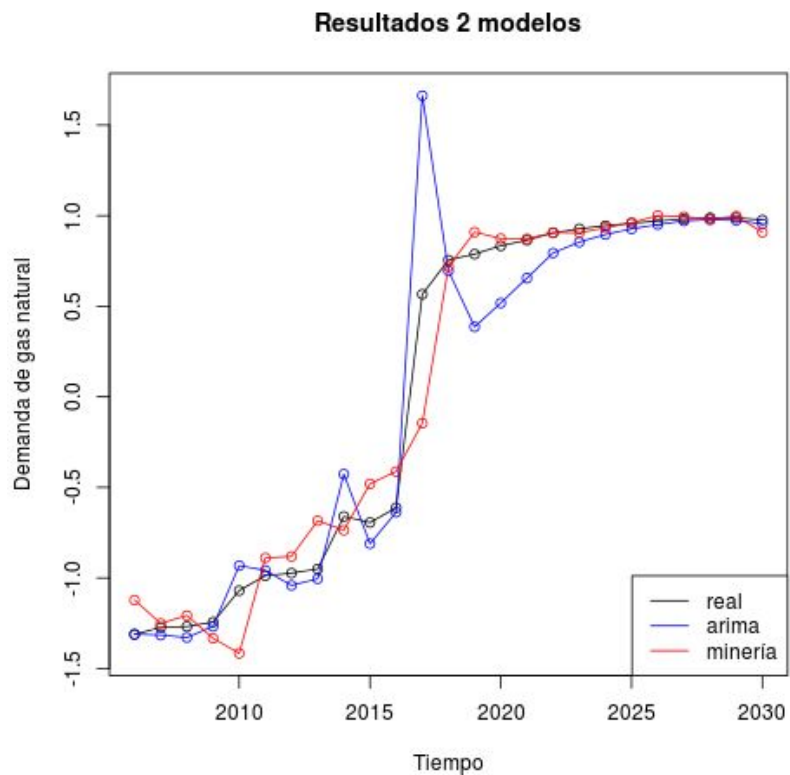


Error TOTAL 2006-2030: 1.63  
i.e. 2X modelo de minería

# Predicciones más allá del período dado



# Los dos modelos



Siguientes pasos

# ¿Qué sigue?

- Refinar modelo
- Comparativa de modelos: pros y contras
- Selección de mejor modelo
- Montar modelos en Power BI



Gracias