Forensic OR: Seeing the Answers, what were the Questions?

John R. Birge

1University of Chicago Booth School of Business (some work joint with Ali Hortaçsu, University of Chicago, and Michael Pavlin, Wilfred Laurier University)

TSL/Loyola/Chicago
In many situations, answers (solutions) may be visible while the questions are not.

Common examples include problems from network industries through transportation and logistics.

Inverse optimization methods can help discover the parameters even when they are only partially identified.

Solution methods resemble those used for recommendation systems and structural estimation.
Example: Commodity Consumption and Prices

- **Observations:**
  - Gasoline prices at A, B, C, D, and Depot
  - Daily production at Depot
  - Daily consumption at A, B, C, and D

- **Unobserved:**
  - Delivery capacity
  - Transportation cost
  - (Storage capacity, Transportation network topology)
Inverse Optimization from Stochastic Solution

- **Observation:**
  - Price doubles at B when demand is high at C

- **Implication:**
  - Following the solution below
  - Transportation capacity of high C and B demand (9)

![Diagram showing effective travel with optimal recourse route.](image-url)
Practical Problem: Prices and Market Power in Electricity Networks

**Background:**
Economics of Two-Stage Electricity Markets
(Veit et al 2006; Sioshansi, Oren, O’Neill 2010; Botterud et al 2011)
Market Power In Electricity Markets
(Cardell, Hitt, Hogan 1996; Jiang, Baldick 2005; Hogan 2012)

**Results rely on assumptions re participant objectives:**
How to obtain what they are?
What are the questions they are trying to answer?

**Example:** Wind Producer Objectives in Midwest ISO
- Two-stages: Day Ahead (DA), Real Time (RT) markets
- Forecast: Wind producers submit a DA production commitment
- Stochastic production: shortfall or surplus made up via RT prices
Practical Problem: Prices and Market Power in Electricity Networks

**Background:**
Economics of Two-Stage Electricity Markets
(Veit et al 2006; Sioshansi, Oren, O’Neill 2010; Botterud et al 2011)
Market Power In Electricity Markets
(Cardell, Hitt, Hogan 1996; Jiang, Baldick 2005; Hogan 2012)

**Results rely on assumptions re participant objectives:**
How to obtain what they are?
What are the questions they are trying to answer?

**Example:** Wind Producer Objectives in Midwest ISO
- Two-stages: Day Ahead (DA), Real Time (RT) markets
- Forecast: Wind producers submit a DA production commitment
- Stochastic production: shortfall or surplus made up via RT prices

**Wind DA Revenues:**

\[ \text{Rev}(Q_{DA}) = Q_{DA} \times P_{DA} + (Q_{RT} - Q_{DA}) \times P_{RT} \]
Practical Problem: Prices and Market Power in Electricity Networks

Background:
Economics of Two-Stage Electricity Markets
(Veit et al 2006; Sioshansi, Oren, O’Neill 2010; Botterud et al 2011)
Market Power In Electricity Markets
(Cardell, Hitt, Hogan 1996; Jiang, Baldick 2005; Hogan 2012)

Results rely on assumptions re participant objectives:
How to obtain what they are?
What are the questions they are trying to answer?

Example: Wind Producer Objectives in Midwest ISO

- Two-stages: Day Ahead (DA), Real Time (RT) markets
- Forecast: Wind producers submit a DA production commitment
- Stochastic production: shortfall or surplus made up via RT prices

Wind DA Revenues:

\[
\text{Rev}(Q_{DA}) = Q_{DA} \times P_{DA} + (Q_{RT} - Q_{DA}) \times P_{RT}
\]

Economic value of intermittent generation depends on forecast quality
(Gowrisankaran, Reynolds, Samano 2011; Skea, Anderson 2008 ... )
The Answers: Data

Midwest ISO 2010 Data:
Forward Premium \( E(P_{DA} - P_{RT}) > $2.00 \)
*But Under-commitment!*

![Graph showing density of wind production with two curves for different times, 23:00 and 00:00.](image)

**Density (wind prod.)**

**Avg. Day Ahead Commitment**
Midwest ISO 2010 Data:
Forward Premium $E(P_{DA} - P_{RT}) > 2.00$

*But Under-commitment!*

Why?
- Bad forecasting?
- Risk aversion?
- Exercise of Market Power?

![Density curve](image)
Midwest ISO 2010 Data:
Forward Premium $E(P_{DA} - P_{RT}) > 2.00$

*But Under-commitment!*

Why?
- Bad forecasting?
- Risk aversion?
- Exercise of Market Power?

Prerequisite for answer:
estimate $E(\partial P_{DA}/\partial Q_{DA})$ and
$E(\partial P_{RT}/\partial Q_{DA})$

Hard with standard econometrics due to endogeneity
Problem: Finding the Right Questions to Fit the Answers

- Ideal solution to endogeneity question:
  An accurate model of price determination process
  - Zonal prices
    uniform price multi-unit auction
    (c.f. Reguant 2012; Hortaçsu, Puller 2008; Wolak 2004)

However, Locational Marginal Pricing dominate North American Markets (e.g. PJM, Midwestern ISO, CAISO, ERCOT)

Prices depend on entire network structure
Net network structure not directly observable

Critical Infrastructure Information Act (2002)

However, Available information:
Midwestern ISO: prices, quantities, bids, active transmission constraints

Can a useful model of the network be inferred from this information?
(Useful for market researchers, participants, designers of CI IA)
Problem: Finding the Right Questions to Fit the Answers

- Ideal solution to endogeneity question:
  
  An accurate model of price determination process

  - Zonal prices
    uniform price multi-unit auction
    (c.f. Reguant 2012; Hortaçsu, Puller 2008; Wolak 2004)

- However, Locational Marginal Pricing dominate North American Markets
  (e.g. PJM, Midwest ISO, CAISO, ERCOT)

  - Prices depend on entire network structure
  - Network structure not directly observable
    Critical Infrastructure Information Act (2002)
Problem: Finding the Right Questions to Fit the Answers

- Ideal solution to endogeneity question:
  An accurate model of price determination process
    - Zonal prices
      uniform price multi-unit auction
      (c.f. Reguant 2012; Hortaçsu, Puller 2008; Wolak 2004)

- However, Locational Marginal Pricing dominate North American Markets
  (e.g. PJM, Midwest ISO, CAISO, ERCOT)
    - Prices depend on entire network structure
    - Network structure not directly observable
      Critical Infrastructure Information Act (2002)

- However, Available information:
  - Midwest ISO: prices, quantities, bids, active transmission constraints

Can a useful model of the network be inferred from this information?
Problem: Finding the Right Questions to Fit the Answers

- Ideal solution to endogeneity question:
  An accurate model of price determination process
  - Zonal prices
    uniform price multi-unit auction
    (c.f. Reguant 2012; Hortaçsu, Puller 2008; Wolak 2004)
- However, Locational Marginal Pricing dominate North American Markets (e.g. PJM, Midwest ISO, CAISO, ERCOT)
  - Prices depend on entire network structure
  - Network structure not directly observable
    Critical Infrastructure Information Act (2002)
- However, Available information:
  - Midwest ISO: prices, quantities, bids, active transmission constraints

Can a useful model of the network be inferred from this information? (Useful for market researchers, participants, designers of CIIA)
Next Steps

- Answers: Locational Marginal Price (LMP) Market: *Linear model*
- Question: the Estimation Problem... *via Inverse Optimization*
- An Algorithm: *A sufficient explanation*
- Application to Data: *Midwest ISO*
Dispatch Model

Relaxation of the unit commitment problem (fixed binary variables):
Dispatch Model

Relaxation of the unit commitment problem (fixed binary variables):

Market Participant $i \in \{1..N\}$

- Produces $x_i$ MWh at an announced cost of $c_i$/MWh
Dispatch Model

Relaxation of the unit commitment problem (fixed binary variables):

Market Participant $i \in \{1..N\}$

- Produces $x_i$ MWh at an announced cost of $c_i$/MWh

Lossless network with links $(i, j)$

- Transmission between $i$ and $j$ of $y_{(i,j)}$ MWh
Dispatch Model

Relaxation of the unit commitment problem (fixed binary variables):

Market Participant \( i \in \{1..N\} \)
- Produces \( x_i \) MWh at an announced cost of \( c_i/\text{MWh} \)

**Lossless** network with links \((i, j)\)
- Transmission between \( i \) and \( j \) of \( y_{i,j} \) MWh
- Topology defined by matrices \( E, A, \) and \( D \)
  - Network Flow Constraints \( Ey = x \)
  - \( R \) Physical Constraints \( Ay = 0 \)
  - \( L \) Transmission Constraints \( Dy \leq d \)
Dispatch Model

Relaxation of the unit commitment problem (fixed binary variables):

Market Participant \( i \in \{1..N\} \)
- Produces \( x_i \) MWh at an announced cost of \( c_i/MWh \)

Lossless network with links \((i,j)\)
- Transmission between \( i \) and \( j \) of \( y_{(i,j)} \) MWh
- Topology defined by matrices \( E, A, \) and \( D \)
  - Network Flow Constraints \( Ey = x \)
  - \( R \) Physical Constraints \( Ay = 0 \)
  - \( L \) Transmission Constraints \( Dy \leq d \)

\[
\begin{align*}
\min c^T x \\
s.t. \quad Ey &= x \\
Ay &= 0 \\
Dy &\leq d \\
u &\geq x \geq l \\
y &\geq 0
\end{align*}
\]
Definition:

1. The Locational Marginal Price (LMP) is the immediate cost of supplying one additional MW of power at a particular node.
2. The LMP is the shadow price of the flow constraint

\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad Ey = x \\
& \quad Ay = 0 \\
& \quad Dy \leq d \\
& \quad u \geq x \geq l \\
& \quad y \geq 0
\end{align*}
\]

\text{dual variables}
\begin{align*}
\pi \\
\sigma \\
\rho \\
\alpha, \gamma
\end{align*}
LMP Example

(Corresponding LP:)

\[
\text{min } 20x_A + 20x_B + 20x_C \\
\text{s.t. } Ey = x \\
2y_{AB} - y_{AC} = 0 \\
y_{BC} \leq 50
\]

(A constraints)

(D constraints)

(Louie, Strunz 2008)

\(C_A: \$20/\text{MWh}\)

\(C_B: \$30/\text{MWh}\)

\(C_C: \$25/\text{MWh}\)
LMP Example

Corresponding LP:

\[
\begin{align*}
\text{min } & 20x_A + 20x_B + 20x_C \\
\text{s.t. } & Ey = x \\
2y_{AB} - y_{AC} = 0 & \quad \text{A constraints} \\
y_{BC} \leq 50 & \quad \text{D constraints} \\
0 \leq x_A \leq 500 & \\
0 \leq x_B \leq 100 & \\
0 \leq x_C \leq 500 & \\
-300 \leq x_D \leq -300 & \\
y \geq 0 &
\end{align*}
\]
LMP Example

Solution:

\[
\begin{align*}
x_A &= 150 & \pi_A &= 20 \\
x_B &= 0 & \pi_B &= 15 \\
x_C &= 150 & \pi_C &= 25 \\
x_D &= -300 & \pi_D &= 25 \\
y_{AB} &= 50 & \rho_{BC} &= -15 \\
y_{AC} &= 100 \\
y_{BC} &= 50 \\
y_{CD} &= 300 \\
\end{align*}
\]
Estimation Problem (Single Sample)

\[ \min c^T x \quad \text{dual variables} \]

s.t.  
\[ Ey = x \]
\[ Ay = 0 \]
\[ Dy \leq d \]
\[ u \geq x \geq l \]
\[ y \geq 0 \]

- Given data:
  - \( c, x, u, l, \pi, \rho \)
- Generate a network model: \( \bar{E}, \bar{A}, \bar{D}, \bar{d} \)
- Explaining shadow prices \( \pi \) and \( \rho \)
General Inverse Optimization

Given:
- a partial specification of an optimization model
- a (partial) specification of an optimal solution

Infer missing model parameters such that:
- Consistency: Known Parameters consistent with optimality
- Simplicity: Missing parameters minimize a norm
General Inverse Optimization

Given:
- a partial specification of an optimization model
- a (partial) specification of an optimal solution

Infer missing model parameters such that:
- Consistency: Known Parameters consistent with optimality
- Simplicity: Missing parameters minimize a norm

Standard form Zhang, Liu 1996,1999 (linear); Ahuja, Orlin 2001 (general):
- Feasible set known
- Opt. solution known
- Cost parameters unknown

Minimize L1 or L∞ norm
Example Inverse Optimization Problem

**Standard form:**
- Feasible set known
- Opt. solution known
- Cost parameters unknown
- Minimize L1 norm

**Electricity Market Problem:**
- Feasible set unknown
- Opt. solution partially known
- Cost parameters known
- Minimize L1 norm

\[
\begin{align*}
\text{min } & \quad c^T x \\
\text{s.t. } & \quad Ey = x \\
& \quad Ay = 0 \\
& \quad Dy \leq d \\
& \quad u \geq x \geq l \\
& \quad y \geq 0
\end{align*}
\]

\[
\begin{align*}
\text{min } & \quad c^T x \\
\text{s.t. } & \quad Ey = x \\
& \quad Ay = 0 \\
& \quad Dy \leq d \\
& \quad u \geq x \geq l \\
& \quad y \geq 0
\end{align*}
\]
Our Inverse Optimization Problem

- Find a “simplest” $\bar{A}, \bar{D}, \bar{d}, \bar{E}$ satisfying optimality conditions
- Minimize 1-Norm
- Regularize $\bar{A}$ s.t. $\sigma_r = 1$
Our Inverse Optimization Problem

- Find a “simplest” $\bar{A}, \bar{D}, \bar{d}, \bar{E}$ satisfying optimality conditions
- Minimize 1-Norm
- Regularize $\bar{A}$ s.t. $\sigma_r = 1$

Resulting Optimization Problem:

$$\min ||\bar{A}||_1 + ||\bar{D}||_1 + ||\bar{d}||_1$$

$$\sum_{r < R} \bar{A}_{r(i,j)} + \sum_{\ell < L} \bar{D}_{\ell(i,j)} \rho(i,j) = \pi_j - \pi_i \quad \forall (i, j) \; \bar{y}_{ij} > 0$$

$$\bar{E}\bar{y} = x$$

$$\bar{A}\bar{y} = 0$$

$$\bar{D}\bar{y} \leq \bar{d}$$
Our Inverse Optimization Problem

- Find a “simplest” $\tilde{A}, \tilde{D}, \tilde{d}, \tilde{E}$ satisfying optimality conditions
- Minimize 1-Norm
- Regularize $\tilde{A}$ s.t. $\sigma_r = 1$

Resulting Optimization Problem:

$$\min ||\tilde{A}||_1 + ||\tilde{D}||_1 + ||\tilde{d}||_1$$

$$\sum_{r<R} \tilde{A}_{r(i,j)} + \sum_{\ell<L} \tilde{D}_{\ell(i,j)} \rho(i,j) = \pi_j - \pi_i$$

$$\forall (i, j) \: \tilde{y}_{ij} > 0$$

$$\tilde{E}\tilde{y} = x$$

$$\tilde{A}\tilde{y} = 0$$

$$\tilde{D}\tilde{y} \leq \tilde{d}$$

2-Step Algorithm:

1. Find $\tilde{E}$ and $\tilde{y}$:
2. Find $\tilde{A}$, $\tilde{D}$, $\tilde{d}$:
Step 1: Determine $\tilde{E}$ and $\tilde{y}$

- Assume no loss $\tilde{E}_{k(ij)} \in \{-1, 0, 1\}$
- Limit to flows between sources and sinks
  
  $\tilde{E}_{i(ij)} = 1$ only if $x_i > 0$,
  
  $\tilde{E}_{j(ij)} = -1$ only if $x_j < 0$,
  
  $\tilde{E}_{k(ij)} = 0$ otherwise

- Minimize requirements on $\tilde{A}$ and $\tilde{D}$ to satisfy
  
  $$\sum_r \tilde{A}_{r(ij)} + \sum_\ell \tilde{D}_{\ell(ij)} \rho(i,j) = \pi_j - \pi_i \quad \forall \tilde{y}_{ij} > 0$$

- By solving:
  
  $$\min \sum_{ij} \tilde{y}_{ij} \cdot \max\{\pi_i - \pi_j, 0\}$$

  s.t. $\tilde{E}\tilde{y} = x$

  $\tilde{y}_{ij} \geq 0$
Step 1: Determine $\bar{E}$ and $\bar{y}$

- Assume no loss $\bar{E}_{k(ij)} \in \{-1, 0, 1\}$
- Limit to paths between sources and sinks
  \[ \bar{E}_{i(ij)} = 1 \text{ only if } x_i > 0, \quad \bar{E}_{j(ij)} = -1 \text{ only if } x_j < 0, \]
  \[ \bar{E}_{k(ij)} = 0 \text{ otherwise} \]
- Minimize requirements on $\bar{A}$ and $\bar{D}$ to satisfy
  \[ \sum_r \bar{A}_{r(i,j)} + \sum_\ell \bar{D}_{\ell(i,j)} \rho(i,j) = \pi_j - \pi_i \quad \forall \bar{y}_{ij} > 0 \]
- By solving:
  \[ \min y_{ij} \cdot \max\{\pi_i - \pi_j, 0\} \]
  \[ \text{s.t. } \bar{E}\bar{y} = \bar{x} \]
  \[ \bar{y}_{ij} \geq 0 \]
Step 2: Determine $\bar{A}$, $\bar{D}$, $\bar{d}$

Minimize sum of 1-norms subject to optimality constraints

Solving:

$$\min \sum_{r,(i,j)} |\bar{A}_{r(ij)}| + \sum_{r,(i,j)} |\bar{D}_{\ell(ij)}| + \sum_{\ell} \bar{d}_\ell$$

s.t. $\bar{A}\bar{y} = 0$

$$\sum_{r} \bar{A}_{r(i,j)} + \sum_{\ell} \bar{D}_{\ell(i,j)} \rho(i,j) = \pi_j - \pi_i \quad \forall \bar{y}_{ij} > 0$$

$$d_\ell \geq 0$$
Multiple Samples

For set of samples 1..S

1. Calculate $\bar{y}^s$ independently
2. Add constraints for each sample

\[
\begin{align*}
\min & \sum_{r,(i,j)} |\bar{A}_{r(i,j)}| + \sum_{r,(i,j)} |\bar{D}_{\ell(i,j)}| + \sum_{\ell} \bar{d}_\ell \\
\text{s.t.} & \quad \bar{A}\bar{y}^s = 0 \quad \forall s \\
& \sum_r \bar{A}_{r(i,j)} + \sum_\ell \bar{D}_{\ell(i,j)} \rho^s_{(i,j)} = \pi^s_j - \pi^s_i \quad \forall \bar{y}^s_{ij} > 0 \\
& \bar{d}_\ell \geq 0
\end{align*}
\]

A and D Constraints:
\[
\frac{1}{3}(\bar{y}_{AD} - 2\bar{y}_{AB}) \leq 50
\]
Observations for each day and hour

- $x, \pi, \rho, c$ vary
- $E, D, A, d$ constant (approximately)

Transmission and generation outages may change $d$

Transmission losses not included

Original “optimization" may be adjusted for other reasons (e.g., frequency, reliability)
Algorithm Observations

- Extension with multiple observations
  - Step 1 performed independently
  - Step 2 single optimization adding all constraints
- Algorithm feasible if rows in $A$ greater number of samples
- Polynomial approximation
  - Step 2: LP standard transformation
  - Each step presents $O(n^2)$ variables
Results: Application to Midwest ISO

2010/01/01 00:00:00
- 1403 Nodes
- 768 Aggregated Nodes
- 772 Active Links
- Transmission bounds per hour
- Imperfect data
- Naive implementation
- 20 ARows $\rightarrow$ 40min

Power Flow in Midwest ISO Network
Comparison Steps

1. **Estimation:** Perform estimation of constraints $\mathcal{K}$ using a set of identifying samples $S_\mathcal{K}$.

2. **In-sample testing:** (IS) Take bids and offers for $s \in S_\mathcal{K}$ determine prices using optimization setup and the estimated constraints.

3. **Out-of-sample testing:** (OS) Take bids and offers for $s \notin S_\mathcal{K}$ where active constraints are a subset of $\mathcal{K}$ and determine prices using optimization model.

4. **Well-out-of-sample testing:** (WOS) Take bids and offers for $s \notin S_\mathcal{K}$ where active constraints are not a subset of $\mathcal{K}$ and determine prices using optimization model.

5. **Assess quality:** To test the model quality, we compare the accuracy to prices generated with two benchmark algorithms. The first benchmark (NC) is a model where we assume that there is no congestion essentially performing a uniform price auction with the observed bids. The second benchmark (UC) produces prices by assuming prices remain unchanged from the first sample to the other samples.
### Test Examples

<table>
<thead>
<tr>
<th>ID</th>
<th>Hour</th>
<th>Active</th>
<th>ID</th>
<th>Hour</th>
<th>Active</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Test1: In sample</strong></td>
<td></td>
<td></td>
<td><strong>Test2: In sample</strong></td>
<td></td>
</tr>
<tr>
<td>IS-1-0</td>
<td>2011-03-26 23:00:00</td>
<td>680</td>
<td>IS-2-0</td>
<td>2011-03-26 23:00:00</td>
<td>680</td>
</tr>
<tr>
<td>IS-1-1</td>
<td>2011-03-26 02:00:00</td>
<td>680</td>
<td>IS-2-1</td>
<td>2011-03-27 00:00:00</td>
<td>680:1473</td>
</tr>
<tr>
<td>IS-2-2</td>
<td>2011-03-26 02:00:00</td>
<td>680</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Out of sample</strong></td>
<td></td>
<td></td>
<td><strong>Out of sample</strong></td>
<td></td>
</tr>
<tr>
<td>OS-1-1</td>
<td>2011-03-26 03:00:00</td>
<td>680</td>
<td>OS-2-1</td>
<td>2011-03-26 03:00:00</td>
<td>680</td>
</tr>
<tr>
<td>OS-1-2</td>
<td>2011-03-27 11:00:00</td>
<td>680</td>
<td>OS-2-2</td>
<td>2011-03-27 01:00:00</td>
<td>680:1473</td>
</tr>
<tr>
<td></td>
<td><strong>Test3: In sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IS-3-0</td>
<td>2011-03-18 12:00:00</td>
<td>100:1526:1530</td>
<td>WOS-0</td>
<td>2011-03-18 21:00:00</td>
<td>1531</td>
</tr>
<tr>
<td>IS-3-1</td>
<td>2011-03-18 13:00:00</td>
<td>100:1526:1530</td>
<td>WOS-1</td>
<td>2011-03-20 22:00:00</td>
<td>100:680:1530</td>
</tr>
<tr>
<td>IS-3-2</td>
<td>2011-03-18 14:00:00</td>
<td>100:1526:1530</td>
<td>WOS-2</td>
<td>2011-03-13 11:00:00</td>
<td>3:307:1257:1496:1504</td>
</tr>
<tr>
<td>IS-3-3</td>
<td>2011-03-18 18:00:00</td>
<td>100:1526:1530</td>
<td>WOS-3</td>
<td>2011-03-20 14:00:00</td>
<td>100:307:1329:1358:1473:1530:1537</td>
</tr>
<tr>
<td></td>
<td><strong>Out of sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS-3-1</td>
<td>2011-03-18 19:00:00</td>
<td>100:1530</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS-3-2</td>
<td>2011-03-18 20:00:00</td>
<td>100:1526:1530</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Test: NC Comparison

<table>
<thead>
<tr>
<th>ID</th>
<th>All locations and participants</th>
<th>Outliers removed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cor-coe std-r-err a-a-err mw a-a-err</td>
<td>cor-coe std-r-err a-a-err mw a-a-err</td>
</tr>
<tr>
<td>IS-1-0</td>
<td>0.893 0.386 5.891 20.187</td>
<td>0.894 0.362 5.741 5.622</td>
</tr>
<tr>
<td>IS-1-1</td>
<td>0.872 0.456 5.578 16.412</td>
<td>0.892 0.383 5.436 5.451</td>
</tr>
<tr>
<td>OS-1-0</td>
<td>0.871 0.460 5.414 16.445</td>
<td>0.897 0.371 5.272 5.409</td>
</tr>
<tr>
<td>OS-1-1</td>
<td>0.947 0.200 12.449 26.323</td>
<td>0.944 0.196 12.338 7.242</td>
</tr>
<tr>
<td>WOS-1-0</td>
<td>0.839 0.563 8.112 21.059</td>
<td>0.839 0.545 8.036 5.857</td>
</tr>
<tr>
<td>WOS-1-1</td>
<td>0.854 0.512 4.512 13.475</td>
<td>0.890 0.402 4.262 4.929</td>
</tr>
<tr>
<td>WOS-1-2</td>
<td>0.470 1.488 15.192 21.887</td>
<td>0.482 1.411 14.958 6.151</td>
</tr>
<tr>
<td>WOS-1-3</td>
<td>0.891 0.389 4.778 12.025</td>
<td>0.925 0.278 4.600 4.216</td>
</tr>
<tr>
<td>IS-2-0</td>
<td>0.893 0.386 5.891 20.057</td>
<td>0.894 0.362 5.741 5.622</td>
</tr>
<tr>
<td>IS-2-1</td>
<td>0.900 0.362 5.992 18.213</td>
<td>0.899 0.345 5.844 5.183</td>
</tr>
<tr>
<td>IS-2-2</td>
<td>0.872 0.456 5.578 16.294</td>
<td>0.892 0.383 5.436 5.451</td>
</tr>
<tr>
<td>OS-2-0</td>
<td>0.871 0.460 5.414 16.331</td>
<td>0.897 0.371 5.272 5.409</td>
</tr>
<tr>
<td>OS-2-1</td>
<td>0.897 0.371 5.506 18.471</td>
<td>0.894 0.357 5.354 5.686</td>
</tr>
<tr>
<td>WOS-2-0</td>
<td>0.839 0.563 8.112 20.922</td>
<td>0.839 0.545 8.036 5.857</td>
</tr>
<tr>
<td>WOS-2-1</td>
<td>0.854 0.512 4.512 13.351</td>
<td>0.890 0.402 4.262 4.929</td>
</tr>
<tr>
<td>WOS-2-2</td>
<td>0.470 1.484 15.221 21.808</td>
<td>0.482 1.407 14.987 6.158</td>
</tr>
<tr>
<td>WOS-2-3</td>
<td>0.891 0.389 4.778 11.905</td>
<td>0.925 0.278 4.600 4.196</td>
</tr>
<tr>
<td>IS-3-0</td>
<td>0.749 0.407 18.133 23.058</td>
<td>0.728 0.354 17.948 5.800</td>
</tr>
<tr>
<td>IS-3-1</td>
<td>0.761 0.389 12.337 22.299</td>
<td>0.737 0.342 12.190 5.312</td>
</tr>
<tr>
<td>IS-3-2</td>
<td>0.719 0.446 8.340 21.076</td>
<td>0.703 0.402 8.197 4.925</td>
</tr>
<tr>
<td>IS-3-3</td>
<td>0.749 0.408 9.486 20.565</td>
<td>0.725 0.373 9.231 4.831</td>
</tr>
<tr>
<td>OS-3-0</td>
<td>0.714 0.455 17.475 21.693</td>
<td>0.686 0.400 17.228 4.956</td>
</tr>
<tr>
<td>OS-3-1</td>
<td>0.812 0.316 11.629 20.557</td>
<td>0.785 0.288 11.486 4.903</td>
</tr>
<tr>
<td>WOS-3-0</td>
<td>0.856 0.247 8.813 23.029</td>
<td>0.863 0.204 8.736 6.179</td>
</tr>
<tr>
<td>WOS-3-1</td>
<td>0.552 0.638 5.306 15.660</td>
<td>0.579 0.627 5.088 5.105</td>
</tr>
<tr>
<td>WOS-3-2</td>
<td>0.614 0.578 15.922 24.198</td>
<td>0.604 0.554 15.686 6.784</td>
</tr>
<tr>
<td>WOS-3-3</td>
<td>0.612 0.574 5.613 14.142</td>
<td>0.632 0.554 5.462 4.337</td>
</tr>
</tbody>
</table>

**Table:** Benchmark 1 NC: no transmission constraints
## Price and MW Predictions

<table>
<thead>
<tr>
<th>ID</th>
<th>All locations and participants</th>
<th>Outliers removed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cor-coef</td>
<td>st-rg-err</td>
</tr>
<tr>
<td>IS-1-0</td>
<td>1.000</td>
<td>0.020</td>
</tr>
<tr>
<td>IS-1-1</td>
<td>0.999</td>
<td>0.042</td>
</tr>
<tr>
<td>OS-1-0</td>
<td>1.000</td>
<td>0.026</td>
</tr>
<tr>
<td>OS-1-1</td>
<td>0.951</td>
<td>7.023</td>
</tr>
<tr>
<td>WOS-1-0</td>
<td>0.723</td>
<td>9.832</td>
</tr>
<tr>
<td>WOS-1-1</td>
<td>0.990</td>
<td>0.334</td>
</tr>
<tr>
<td>WOS-1-2</td>
<td>0.349</td>
<td>16.542</td>
</tr>
<tr>
<td>WOS-1-3</td>
<td>0.974</td>
<td>1.048</td>
</tr>
<tr>
<td>IS-2-0</td>
<td>1.000</td>
<td>0.010</td>
</tr>
<tr>
<td>IS-2-1</td>
<td>0.470</td>
<td>107.298</td>
</tr>
<tr>
<td>IS-2-2</td>
<td>0.993</td>
<td>0.311</td>
</tr>
<tr>
<td>OS-2-0</td>
<td>0.842</td>
<td>11.519</td>
</tr>
<tr>
<td>OS-2-1</td>
<td>0.456</td>
<td>109.181</td>
</tr>
<tr>
<td>WOS-2-0</td>
<td>0.723</td>
<td>9.830</td>
</tr>
<tr>
<td>WOS-2-1</td>
<td>0.988</td>
<td>0.338</td>
</tr>
<tr>
<td>WOS-2-2</td>
<td>0.350</td>
<td>16.432</td>
</tr>
<tr>
<td>WOS-2-3</td>
<td>0.974</td>
<td>0.926</td>
</tr>
<tr>
<td>IS-3-0</td>
<td>0.416</td>
<td>687.254</td>
</tr>
<tr>
<td>IS-3-1</td>
<td>0.567</td>
<td>180.706</td>
</tr>
<tr>
<td>IS-3-2</td>
<td>0.399</td>
<td>669.760</td>
</tr>
<tr>
<td>IS-3-3</td>
<td>0.384</td>
<td>4119.915</td>
</tr>
<tr>
<td>OS-3-0</td>
<td>0.157</td>
<td>26105.651</td>
</tr>
<tr>
<td>OS-3-1</td>
<td>0.210</td>
<td>28562.426</td>
</tr>
<tr>
<td>WOS-3-0</td>
<td>0.140</td>
<td>9407.944</td>
</tr>
<tr>
<td>WOS-3-1</td>
<td>0.744</td>
<td>21.780</td>
</tr>
<tr>
<td>WOS-3-2</td>
<td>0.314</td>
<td>205.694</td>
</tr>
<tr>
<td>WOS-3-3</td>
<td>0.581</td>
<td>105.566</td>
</tr>
</tbody>
</table>
LMP Predictions

(a) In Sample

(b) Out of Sample

Figure: Predicted vs. True LMPs, estimation using Test 2 with two active constraint
Additional Examples

- **Macro-economic models**
  - Prices observed in different regions
  - Purchase quantities observed
  - Equilibrium model set up as potential optimization
  - Unknown transportation routes and costs to discover

- **Supply chain interactions**
  - Prices and quantities observed
  - Unobserved relationships between suppliers and customers
  - Discover relationships and transactions
Future Directions

- Discussed modelling price determination process in LMP based Electricity Markets
  - Inverse optimization based formulation/algorithm consistent with dual interpretation of LMPs

Next Steps
- Predicting market characteristics: *price response, congestion costs*...
- Structural estimation: *model participant decision making*
- Solution quality: *solution robustness to data imperfections*
- Econometrics of general competitive markets: *extend to linear market models*
Conclusions

- Sometimes we have the questions but need to know the questions to predict future outcomes.
- Can use inverse optimization to find the right questions.
  - Needed to determine objective of agents.
  - Applies to many markets with price and quantity observations but not constraints or costs.
  - Often had a difficulty from bilinear form with constraint and unknown variable values.
- Solution method.
  - Two-step process.
  - Determine consistent primal variables first.
  - Choose constraint coefficients with minimum 1-norm.
- Results.
  - Possible to discover simple network configurations.
  - Reasonable results with multiple data observations.
  - Possible inconsistencies from unknown parameter changes (or may be a way to detect some form of infiltration).
Thank you! Questions?