Horse Race Models: A Comprehensive Guide

What is a Horse Race Model?

A **horse race model** is a regression specification where you include **multiple competing explanations** (variables) simultaneously to see which one "wins" - i.e., which remains statistically significant when they compete for explanatory power.

The Metaphor

Like horses racing on a track - you put multiple variables in the same model and see which ones are still "running strong" (statistically significant) after controlling for the others.

Your Application: Gas Prices and Voter Turnout

The Question

Does **current** gas price affect turnout, or is it really just **lagged** gas prices (persistent trends)?

The Horse Race Model

`r turnout ~ current_gas + lagged_gas + controls`

Your Results

| Variable | Coefficient | P-value | Interpretation | |----------|-------------|---------|----------------| | **Current gas** | 0.079 | 0.001 | âœ… **Winner** - Highly significant | | **Lagged gas** | 0.017 | 0.61 | âŒ Not significant |

Conclusion: Current gas prices "win the race" - they remain significant while lagged prices don't. This proves your effect is real and contemporaneous, not just picking up persistent trends.

Four Possible Outcomes in a Horse Race

Outcome 1: Variable A Wins (Your Result)

- Variable A: **Significant**
- Variable B: Not significant
- **Interpretation:** A is the true driver; B was spuriously correlated

Outcome 2: Variable B Wins

- Variable A: Not significant
- Variable B: **Significant**
- **Interpretation:** B is the true driver; A was spuriously correlated

Outcome 3: Both Win

- Variable A: **Significant**
- Variable B: **Significant**
- **Interpretation:** Both matter independently (e.g., cumulative effects)

Outcome 4: Neither Wins

- Variable A: Not significant
- Variable B: Not significant
- **Interpretation:** High multicollinearity - they're so correlated they cancel each other out

---

Classic Examples in Economics

Example 1: Returns to Education

Question: Does education increase wages, or is it really just coming from rich families?

`r wages ~ years_education + parent_income + controls`

Possible findings: - If education wins: Education has causal effect on wages - If parent income wins: Education is just a proxy for family wealth - If both win: Both education and family background matter

---

Example 2: Campaign Spending

Question: Does incumbent spending help win votes, or is it just an arms race where only relative spending matters?

`r vote_share ~ incumbent_spending + challenger_spending + controls`

Possible findings: - If incumbent spending wins: More spending = more votes - If neither wins separately but difference matters: It's about relative spending - If both negative: Spending signals weakness (quality candidates don't need to spend)

Example 3: Gubernatorial Incumbent Voting

Question: Which timing of gas prices actually matters for incumbent support?

`r incumbent_vote ~ current_gas + past_gas + future_gas + controls`

In your analysis: - Current gas: Small, not significant - Past gas: Larger in isolation, but disappears in horse race - Future gas: Not significant - **Conclusion:** Gas prices don't meaningfully affect gubernatorial voting

---

Key Insights from Horse Race Models

What They Test

Horse race models determine whether an effect is:

1. **Direct/Causal** â†' Variable stays significant when competitors included
2. **Spurious** â†' Variable becomes insignificant when controlling for the real cause
3. **Shared/Confounded** â†' Both lose significance due to multicollinearity

Why They're Useful

1. **Rule out omitted variable bias** - Include the potential omitted variable directly
2. **Test competing theories** - Let the data choose between alternative explanations
3. **Build credibility** - Show your main effect survives tough competition

---

Your Specific Case: Addressing Serial Correlation

The Problem

Gas prices are **serially correlated** - if prices are high this year, they were probably high last year too.

The Concern

When you estimate: `r turnout ~ current_gas + controls`

Lagged gas prices are an **omitted variable** that's correlated with: 1. Current gas prices (serial correlation) 2. Current turnout (persistent state-level trends)

This creates **omitted variable bias**.

The Solution: Horse Race

`r turnout ~ current_gas + lagged_gas + controls`

By including lagged gas prices: - You control for the omitted variable - The lagged effect goes to zero (it was spurious) - The current effect remains strong (it's real) - You isolate the true contemporaneous causal effect

---

What NOT to Confuse With Horse Race Models

NOT the Same As:

1. **Stepwise Regression**
2. Automatically adding/removing variables based on significance

3. Generally discouraged (p-hacking, data mining)

4. **Variable Selection**
5. Choosing which variables to include a priori
6. Based on theory, not competition
7. **Specification Search**
8. Trying many different model specifications
9. Problematic for inference

Horse Race Is:

- **Theory-driven**: You have specific competing explanations to test
- **Transparent**: You show both the simple model and the horse race
- **Hypothesis testing**: Testing whether one explanation dominates another

---

How to Present in Your Paper

Main Results Section

Present the **simple model** first (current gas price only): - Coefficient = 0.056, p < 0.001 - This is your main finding - Clean, easy to interpret - Focus on substantive interpretation

Robustness Checks Section

``` We conduct several robustness tests to ensure our results are not driven by spurious correlations:

1. Placebo tests: Future gas prices show no effect (p=0.19), confirming no reverse causality or anticipation effects.

2. Serial correlation: Lagged gas prices appear significant in isolation (coefficient=0.072, p=0.004), raising concerns about persistent trends. However, when we include both current and lagged gas prices in a horse race specification, lagged prices become insignificant (coefficient=0.017, p=0.61) while the current effect strengthens (coefficient=0.079, p=0.001). This confirms our results reflect contemporaneous effects rather than omitted variable bias from persistent state-level trends. ```

### Why This Order?

1. **Simpler model first** - Easier for readers to understand
2. **Build credibility** - Show you anticipated concerns
3. **Standard practice** - What reviewers expect

---

# Summary: Your Three Key Findings

### 1. Voter Turnout âœ…

- **Main effect:** Gas prices increase turnout (0.056, p < 0.001)
- **Horse race:** Current prices matter (0.079, p = 0.001), lagged don't (p = 0.61)
- **Conclusion:** Real contemporaneous effect, not persistent trends

### 2. Presidential Incumbent Voting âœ…

- **Main effect:** Gas prices hurt incumbent presidents (significant, negative)
- **Conclusion:** Classic economic voting

### 3. Gubernatorial Incumbent Voting âš ï¸

- **Main effect:** Null (-0.02, p = 0.69)
- **Placebos:** Past effects larger than current (concerning, but effect is null anyway)
- **Conclusion:** Voters don't hold governors accountable for gas prices

---

# The Big Picture

Your paper tells a coherent story:

1. **Gas prices mobilize voters** â†’ Higher turnout when prices are salient
2. **Voters punish presidents for gas prices** â†’ National economic accountability
3. **But NOT governors** â†’ Voters correctly attribute gas prices to national/global factors, not state government
```

This is a **strong contribution** showing that: - Economic voting depends on attribution of responsibility - Voters are sophisticated enough to distinguish state vs. national issues - Gas prices have broader political effects beyond just hurting incumbents (mobilization!)

Technical Note: When to Use Horse Race Models

Good Use Cases:

âœ… Testing specific competing theories with clear predictions âœ… Addressing concerns about omitted variable bias âœ… When variables are correlated but theoretically distinct âœ… Robustness checks in published research

Bad Use Cases:

âŒ Fishing for significance among many variables âŒ When variables measure the same underlying construct âŒ As a substitute for proper causal identification strategy âŒ When multicollinearity is extreme (VIF > 10)

Your Use Case: âœ… Perfect Application

You had a specific concern (serial correlation), a clear alternative explanation (persistent trends), and you tested it directly. This is exactly what horse race models are designed for.