Occasionally Misspecified

JJ Forneron, Boston University

October 2023

Motivation

- Want to estimate a model from a sample of data:
  - good fit for most of the sample
  - grossly misspecified for a few observations

Motivation

- Want to estimate a model from a sample of data:
  - good fit for most of the sample
  - grossly misspecified for a few observations
- Can be due to:
  - temporary change in policy rule
  - imperfect match between merged datasets
  - generated regressors/outcomes with some large mispredictions

Motivation

- Want to estimate a model from a sample of data:
  - good fit for most of the sample
  - grossly misspecified for a few observations
- Can be due to:
  - temporary change in policy rule
  - imperfect match between merged datasets
  - generated regressors/outcomes with some large mispredictions
- Solution: **robust estimation**
  + less sensitive to outliers

Motivation

- Want to estimate a model from a sample of data:
  - good fit for most of the sample
  - grossly misspecified for a few observations
- Can be due to:
  - temporary change in policy rule
  - imperfect match between merged datasets
  - generated regressors/outcomes with some large mispredictions
- Solution: **robust estimation**
  - $+$ less sensitive to outliers
  - $-$ **inconsistent/biased** unless **data symmetric** + **symmetrically contaminated**

Motivation

- Want to estimate a model from a sample of data:
  - good fit for most of the sample
  - grossly misspecified for a few observations
- Can be due to:
  - temporary change in policy rule
  - imperfect match between merged datasets
  - generated regressors/outcomes with some large mispredictions
- Solution: **robust estimation**
  - $+$ less sensitive to outliers
  - $-$ **inconsistent/biased** unless **data symmetric** $+$ **symmetrically contaminated**
- Look for an estimator that is:

Motivation

- Want to estimate a model from a sample of data:
  - good fit for most of the sample
  - grossly misspecified for a few observations
- Can be due to:
  - temporary change in policy rule
  - imperfect match between merged datasets
  - generated regressors/outcomes with some large mispredictions
- Solution: **robust estimation**
  - \+ less sensitive to outliers
  - − **inconsistent/biased** unless **data symmetric** + **symmetrically contaminated**
- Look for an estimator that is:
  1. robust to small amounts of large contamination

Motivation

- Want to estimate a model from a sample of data:
  - good fit for most of the sample
  - grossly misspecified for a few observations
- Can be due to:
  - temporary change in policy rule
  - imperfect match between merged datasets
  - generated regressors/outcomes with some large mispredictions
- Solution: **robust estimation**
  + less sensitive to outliers
  − **inconsistent/biased** unless **data symmetric** + **symmetrically contaminated**
- Look for an estimator that is:
  1. robust to small amounts of large contamination
  2. asymptotically unbiased

Motivation

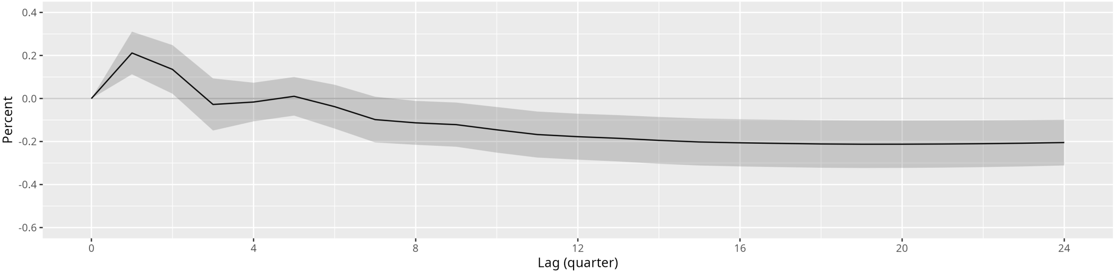
- Want to estimate a model from a sample of data:
  - good fit for most of the sample
  - grossly misspecified for a few observations
- Can be due to:
  - temporary change in policy rule
  - imperfect match between merged datasets
  - generated regressors/outcomes with some large mispredictions
- Solution: **robust estimation**
  - $+$ less sensitive to outliers
  - $-$ **inconsistent/biased** unless **data symmetric** $+$ **symmetrically contaminated**
- Look for an estimator that is:
  1. robust to small amounts of large contamination
  2. asymptotically unbiased
  3. not too hard to compute

Motivating Example: Price Puzzle

- Estimand: effect of monetary shock on inflation
- Method: recursive VAR (OLS)
  - Variables: Interest Rates R_t, Inflation π_t, Unemployment: u_t
  - Specification: 4 lags
  - Data: US from 1960Q1 to 2000Q4
    (same as Stock and Watson, 2001)

Motivating Example: Price Puzzle

- Estimand: effect of monetary shock on inflation
- Method: recursive VAR (OLS)
  - Variables: Interest Rates R_t, Inflation π_t, Unemployment: u_t
  - Specification: 4 lags
  - Data: US from 1960Q1 to 2000Q4
    (same as Stock and Watson, 2001)
- Results: when $R_t \nearrow$, $\pi_t \nearrow$



Price Puzzle: Background

- Initially Sims (1992), termed "Price Puzzle" by Eichenbaum (1992)

Price Puzzle: Background

- Initially Sims (1992), termed "Price Puzzle" by Eichenbaum (1992)
- Several explanations, resolutions:
  - omitted variables: e.g. commodity prices (Sims, 1992), factors (Bernanke et al., 2005)
  - structural breaks: e.g (Castelnuovo and Surico, 2010)
  - ...
  - meta-analysis: Rusnák et al. (2013) look at 1000 regressions ($\gg T$)

Price Puzzle: Background

- Initially Sims (1992), termed "Price Puzzle" by Eichenbaum (1992)
- Several explanations, resolutions:
  - omitted variables: e.g. commodity prices (Sims, 1992), factors (Bernanke et al., 2005)
  - structural breaks: e.g (Castelnuovo and Surico, 2010)
  - ...
  - meta-analysis: Rusnák et al. (2013) look at 1000 regressions ($\gg T$)
- The initial impact at Lag $= 1$ measured by β_1 in:

$$\pi_t = \beta_0 + \beta_1 R_{t-1} + \beta_2 u_{t-1} + \beta_3 \pi_{t-1} + \cdots + \beta_{10} R_{t-4} + \beta_{11} u_{t-4} + \beta_{12} \pi_{t-4} + e_{\pi,t}.$$

estimated by OLS

Price Puzzle: Background

- Initially Sims (1992), termed "Price Puzzle" by Eichenbaum (1992)
- Several explanations, resolutions:
  - omitted variables: e.g. commodity prices (Sims, 1992), factors (Bernanke et al., 2005)
  - structural breaks: e.g (Castelnuovo and Surico, 2010)
  - ...
  - meta-analysis: Rusnák et al. (2013) look at 1000 regressions ($\gg T$)
- The initial impact at Lag $= 1$ measured by β_1 in:

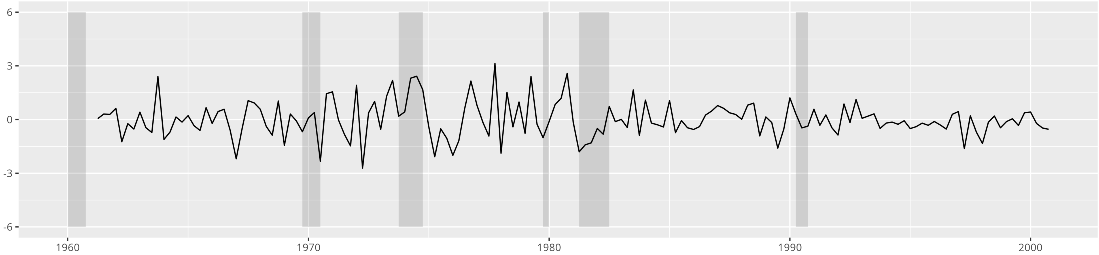
$$\pi_t = \beta_0 + \beta_1 R_{t-1} + \beta_2 u_{t-1} + \beta_3 \pi_{t-1} + \cdots + \beta_{10} R_{t-4} + \beta_{11} u_{t-4} + \beta_{12} \pi_{t-4} + e_{\pi,t}.$$

  estimated by OLS
- Let's look at some regression diagnostics

Price Puzzle: Diagnostics

$$\pi_t = \beta_0 + \beta_1 R_{t-1} + \beta_2 u_{t-1} + \beta_3 \pi_{t-1} + \cdots + \beta_{10} R_{t-4} + \beta_{11} u_{t-4} + \beta_{12} \pi_{t-4} + e_{\pi,t}.$$

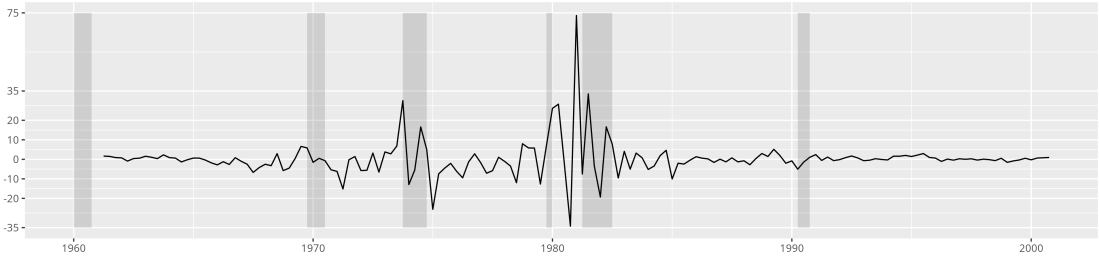
- Residuals $\hat{e}_{\pi,t}$ (standardized)



- Skewness: 0.36, Kurtosis: 3.78

Price Puzzle: Diagnostics

- Contributions to $\hat{\beta}_{1n}$ – based on $(X'X/n)^{-1}x_t y_t$ (avg $= \hat{\beta}_n$)



- 1981Q1: $75/n \simeq 0.47$ vs. $\hat{\beta}_1 = 0.21$, 3.5 standard errors
- Skewness: 3.24, Kurtosis: 27.81

Price Puzzle: Concerns

- 1979-1982: Federal Reserve no longer sets R_t directly
  - Non-borrowed reserves targeting (misspecification)
  - Increased volatility in R_t (leverage)

$\Rightarrow$ **Leveraged outliers**

Price Puzzle: Concerns

- 1979-1982: Federal Reserve no longer sets R_t directly
  - Non-borrowed reserves targeting (misspecification)
  - Increased volatility in R_t (leverage)

$\Rightarrow$ **Leveraged outliers**

- Robust estimation & inference desirable
  - $+$ Reduce the influence of outliers
  - $-$ Biased/Inconsistent for asymmetric data

Price Puzzle: Concerns

- 1979-1982: Federal Reserve no longer sets R_t directly
  - Non-borrowed reserves targeting (misspecification)
  - Increased volatility in R_t (leverage)

$\Rightarrow$ **Leveraged outliers**

- Robust estimation & inference desirable
  - $+$ Reduce the influence of outliers
  - $-$ Biased/Inconsistent for asymmetric data
  - $-$ Robust M-estimators **not robust to leverage** (Hamilton, 1992)

Literature

- Robust M-estimation:

$$\min_{\theta} \sum_{t=1}^{n} \psi(y_t - x_t'\theta),$$

  ψ non-quadratic: LAD, Huber loss, trimming, etc.

Literature

- Robust M-estimation:

$$\min_{\theta} \sum_{t=1}^{n} \psi(y_t - x_t'\theta),$$

  ψ non-quadratic: LAD, Huber loss, trimming, etc.
- Classic asymptotic results:
  - Huber (1964): $F_\varepsilon = (1-\varepsilon)\Phi + \varepsilon G$, symmetric

Literature

- Robust M-estimation:
$$\min_{\theta} \sum_{t=1}^{n} \psi(y_t - x_t'\theta),$$
  ψ non-quadratic: LAD, Huber loss, trimming, etc.
- Classic asymptotic results:
  - Huber (1964): $F_\varepsilon = (1-\varepsilon)\Phi + \varepsilon G$, symmetric
  - Jaeckel (1971):
$$F_{\varepsilon_n} = (1-\varepsilon_n)\Phi + \varepsilon_n G,$$
    G asymmetric, $\varepsilon_n = O(n^{-1/2}) \Rightarrow$ asymptotic bias $= O(n^{-1/2})$
  - Hampel (1974): Influence Curve for local asymptotics

Literature

- Robust M-estimation:

$$\min_{\theta} \sum_{t=1}^{n} \psi(y_t - x_t'\theta),$$

  ψ non-quadratic: LAD, Huber loss, trimming, etc.
- Classic asymptotic results:
  - Huber (1964): $F_\varepsilon = (1-\varepsilon)\Phi + \varepsilon G$, symmetric
  - Jaeckel (1971):

$$F_{\varepsilon_n} = (1-\varepsilon_n)\Phi + \varepsilon_n G,$$

    G asymmetric, $\varepsilon_n = O(n^{-1/2}) \Rightarrow$ asymptotic bias $= O(n^{-1/2})$
  - Hampel (1974): Influence Curve for local asymptotics
  - F_0 asymmetric $\Rightarrow$ inconsistency
    Carroll and Welsh (1988), Newey and Steigerwald (1997),...

Literature

- Robust M-estimation:
$$\min_\theta \sum_{t=1}^{n} \psi(y_t - x_t'\theta),$$
ψ non-quadratic: LAD, Huber loss, trimming, etc.
- Classic asymptotic results:
  - Huber (1964): $F_\varepsilon = (1-\varepsilon)\Phi + \varepsilon G$, symmetric
  - Jaeckel (1971):
  $$F_{\varepsilon_n} = (1-\varepsilon_n)\Phi + \varepsilon_n G,$$
  G asymmetric, $\varepsilon_n = O(n^{-1/2}) \Rightarrow$ asymptotic bias $= O(n^{-1/2})$
  - Hampel (1974): Influence Curve for local asymptotics
  - F_0 asymmetric $\Rightarrow$ inconsistency
  Carroll and Welsh (1988), Newey and Steigerwald (1997),...
  - Cantoni and Ronchetti (2001): GLM, analytical bias correction (parametric)

Literature

- Robust M-estimation:

$$\min_{\theta} \sum_{t=1}^{n} \psi(y_t - x_t'\theta),$$

  ψ non-quadratic: LAD, Huber loss, trimming, etc.
- Classic asymptotic results:
  - Huber (1964): $F_\varepsilon = (1-\varepsilon)\Phi + \varepsilon G$, symmetric
  - Jaeckel (1971):

$$F_{\varepsilon_n} = (1-\varepsilon_n)\Phi + \varepsilon_n G,$$

    G asymmetric, $\varepsilon_n = O(n^{-1/2}) \Rightarrow$ asymptotic bias $= O(n^{-1/2})$
  - Hampel (1974): Influence Curve for local asymptotics
  - F_0 asymmetric $\Rightarrow$ inconsistency
    Carroll and Welsh (1988), Newey and Steigerwald (1997),...
  - Cantoni and Ronchetti (2001): GLM, analytical bias correction (parametric)
- Local asymptotics imply $\hat{\theta}_n$ consistent

Literature

- More recent, finite-sample: Median-of-Means
  - K-subsamples, K means, return median
  - robust up to $n_o \leq K/2 - 1$ outliers
  - cv. rate depends on n/K
  - e.g. Lecué and Lerasle (2020), Laforgue et al. (2021)

Literature

- More recent, finite-sample: Median-of-Means
  - K-subsamples, K means, return median
  - robust up to $n_o \leq K/2 - 1$ outliers
  - cv. rate depends on n/K
  - e.g. Lecué and Lerasle (2020), Laforgue et al. (2021)
- Bias can be bounded but not tractable:

$$|\text{median}(X) - \mathbb{E}_P(X)| \leq \sigma_P(X)$$

Literature

- More recent, finite-sample: Median-of-Means
  - K-subsamples, K means, return median
  - robust up to $n_o \leq K/2 - 1$ outliers
  - cv. rate depends on n/K
  - e.g. Lecué and Lerasle (2020), Laforgue et al. (2021)
- Bias can be bounded but not tractable:

$$|\text{median}(X) - \mathbb{E}_P(X)| \leq \sigma_P(X)$$

- $n_o \to \infty$ requires $K \to \infty$
- bias of order: $\sqrt{K/n}$

Setting: Sample

- GMM estimation: $\mathbb{E}_P[g(z_t;\theta)] = 0 \Leftrightarrow \theta = \theta_0$
- Data $(z_1, \ldots, z_n)$ with $n = n_P + n_o$

Setting: Sample

- GMM estimation: $\mathbb{E}_P[g(z_t;\theta)] = 0 \Leftrightarrow \theta = \theta_0$
- Data $(z_1, \ldots, z_n)$ with $n = n_P + n_o$
- n_P **good observations**: $t = 1, \ldots, n_P$ s.t. $z_t \sim P$
  - iid or
  - strictly stationary, β-mixing with $\beta_m \leq a \exp(-bm)$ for $0 < a, b < \infty$

Setting: Sample

- GMM estimation: $\mathbb{E}_P[g(z_t;\theta)] = 0 \Leftrightarrow \theta = \theta_0$
- Data $(z_1, \ldots, z_n)$ with $n = n_P + n_o$
- n_P **good observations**: $t = 1, \ldots, n_P$ s.t. $z_t \sim P$
  - iid or
  - strictly stationary, β-mixing with $\beta_m \leq a \exp(-bm)$ for $0 < a, b < \infty$
- n_o **outliers**: for $0 \leq A, \alpha < \infty$ and $t = n_P + 1, \ldots, n$

$$z_t \in \mathcal{O}_n = \{z, \sup_{\theta \in \Theta} \|g(z;\theta)\| \leq An^{\alpha}\}$$

Setting: Sample

- GMM estimation: $\mathbb{E}_P[g(z_t;\theta)] = 0 \Leftrightarrow \theta = \theta_0$
- Data $(z_1, \ldots, z_n)$ with $n = n_P + n_o$
- n_P **good observations**: $t = 1, \ldots, n_P$ s.t. $z_t \sim P$
    - iid or
    - strictly stationary, β-mixing with $\beta_m \leq a \exp(-bm)$ for $0 < a, b < \infty$
- n_o **outliers**: for $0 \leq A, \alpha < \infty$ and $t = n_P + 1, \ldots, n$

$$z_t \in \mathcal{O}_n = \{z, \sup_{\theta \in \Theta} \|g(z;\theta)\| \leq An^\alpha\}$$

- Sample mean $\overline{g}_n(\theta)$ can be
    - asymptotically biased if $n_o n^\alpha / \sqrt{n} = O(1)$
    - inconsistent if $n_o n^\alpha / n = O(1)$
    - divergent if $n_o n^\alpha / n \to \infty$

Setting: Estimator

1. **Moments:** find $\hat{\psi}_n(\theta;\nu) = (\hat{\mu}_n, \hat{\Sigma}_n)$ minimizing:

$$Q_n(\psi;\theta) = \frac{\nu + p}{n} \sum_{t=1}^{n} \log\left(1 + \frac{\|g(z_t;\theta) - \mu\|^2_{\Sigma^{-1}}}{\nu}\right) + \log|\Sigma| + \frac{\kappa_1}{\nu}\|\mu\|^2_{\Sigma^{-1}} + \frac{\kappa_2}{\nu}\mathrm{trace}(\Sigma)$$

for $0 < \nu, \kappa_1, \kappa_2 < \infty$ over

$$\Psi = \{(\mu, \Sigma),\ \mu \in \mathbb{R}^p, 0 < s_0 \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq +\infty\}$$

Setting: Estimator

1. **Moments:** find $\hat{\psi}_n(\theta;\nu) = (\hat{\mu}_n, \hat{\Sigma}_n)$ minimizing:

$$Q_n(\psi;\theta) = \frac{\nu+p}{n}\sum_{t=1}^{n}\log\left(1+\frac{\|g(z_t;\theta)-\mu\|^2_{\Sigma^{-1}}}{\nu}\right)+\log|\Sigma|+\frac{\kappa_1}{\nu}\|\mu\|^2_{\Sigma^{-1}}+\frac{\kappa_2}{\nu}\mathrm{trace}(\Sigma)$$

   for $0 < \nu, \kappa_1, \kappa_2 < \infty$ over

$$\Psi = \{(\mu,\Sigma),\ \mu\in\mathbb{R}^p, 0 < s_0 \le \lambda_{\min}(\Sigma) \le \lambda_{\max}(\Sigma) \le +\infty\}$$

2. **Correction:**

$$\tilde{\mu}_n(\theta) = 2\hat{\mu}_n(\theta;\nu) - \hat{\mu}_n(\theta;\nu/2)$$

Setting: Estimator

1. **Moments:** find $\hat{\psi}_n(\theta;\nu) = (\hat{\mu}_n, \hat{\Sigma}_n)$ minimizing:

$$Q_n(\psi;\theta) = \frac{\nu+p}{n}\sum_{t=1}^{n}\log\left(1+\frac{\|g(z_t;\theta)-\mu\|^2_{\Sigma^{-1}}}{\nu}\right)+\log|\Sigma|+\frac{\kappa_1}{\nu}\|\mu\|^2_{\Sigma^{-1}}+\frac{\kappa_2}{\nu}\text{trace}(\Sigma)$$

   for $0<\nu,\kappa_1,\kappa_2<\infty$ over

$$\Psi = \{(\mu,\Sigma),\ \mu\in\mathbb{R}^p, 0 < s_0 \le \lambda_{\min}(\Sigma) \le \lambda_{\max}(\Sigma) \le +\infty\}$$

2. **Correction:**

$$\tilde{\mu}_n(\theta) = 2\hat{\mu}_n(\theta;\nu) - \hat{\mu}_n(\theta;\nu/2)$$

3. **Estimation:** find $\tilde{\theta}_n$ such that

$$\|\tilde{\mu}_n(\tilde{\theta}_n)\|^2_{W_n} \le \inf_{\theta\in\Theta}\|\tilde{\mu}_n(\theta)\|^2_{W_n} + o_p(n^{-1})$$

Tuning parameters, Some Properties

- $0 < \nu < \infty$: controls robustness,
  - $\nu = \infty$: $\hat{\mu}_n(\theta; \infty) = \overline{g}_n(\theta)$ sample mean, $\hat{\Sigma}_n(\theta; \infty)$ sample covariance
  - $\nu < \infty$: $\hat{\mu}_n(\theta; \nu)$, $\hat{\Sigma}_n(\theta; \nu)$ biased

Tuning parameters, Some Properties

- $0 < \nu < \infty$: controls robustness,
  - $\nu = \infty$: $\hat{\mu}_n(\theta; \infty) = \overline{g}_n(\theta)$ sample mean, $\hat{\Sigma}_n(\theta; \infty)$ sample covariance
  - $\nu < \infty$: $\hat{\mu}_n(\theta; \nu)$, $\hat{\Sigma}_n(\theta; \nu)$ biased
- $\kappa_1 = 0$
  - $\partial_\mu Q_n(\psi; \theta) = 0$ when $\|\mu\| = \infty$, for any Σ
  - can be numerically unstable

Tuning parameters, Some Properties

- $0 < \nu < \infty$: controls robustness,
  - $\nu = \infty$: $\hat{\mu}_n(\theta;\infty) = \overline{g}_n(\theta)$ sample mean, $\hat{\Sigma}_n(\theta;\infty)$ sample covariance
  - $\nu < \infty$: $\hat{\mu}_n(\theta;\nu)$, $\hat{\Sigma}_n(\theta;\nu)$ biased
- $\kappa_1 = 0$
  - $\partial_\mu Q_n(\psi;\theta) = 0$ when $\|\mu\| = \infty$, for any Σ
  - can be numerically unstable
- Weighted average representation:

$$\hat{\mu}_n(\theta;\nu) = \sum_t \omega_t(\theta;\nu) g(z_t;\theta), \quad \tilde{\mu}_n(\theta;\nu) = \sum_t \tilde{\omega}_t(\theta;\nu) g(z_t;\theta)$$

where $0 \le \omega_t$, $\sum_t \omega_t \le 1$, $\tilde{\omega}_t = 2\omega_t(\nu) - \omega_t(\nu/2)$

$\Rightarrow$ Robust-LS is weighted-LS with weights $\tilde{w}_t$

Some intuition

- Simplified estimator: $\theta_0 = \mathbb{E}_P(z_t)$

$$\hat{\mu}_n(\nu) = \frac{1}{n}\sum_{t=1}^{n} \frac{z_t}{1 + |z_t|^2/\nu}$$

Some intuition

- Simplified estimator: $\theta_0 = \mathbb{E}_P(z_t)$

$$\hat{\mu}_n(\nu) = \frac{1}{n}\sum_{t=1}^{n} \frac{z_t}{1 + |z_t|^2/\nu}$$

- Finite-Sample: a) $\sup_z |z|/(1 + |z|^2/\nu) \leq \sqrt{\nu}/2$, b) Bernstein's inequality:

$$\mathbb{P}\left(\sup_{z_t \in \mathcal{O}_n, t > n_P} |\hat{\mu}_n(\nu) - \mu(\nu)| \geq \frac{\sqrt{\nu} n_o}{n} + \frac{n_P}{n}\frac{x}{\sqrt{n_P}}\right) \leq 2\exp\left(-\frac{x^2}{2}\sigma_\nu^2 + 2/3\sqrt{\nu/n_P}x\right)$$

when $z_t \sim P$, $1 \leq t \leq n_P$, are iid. $\sigma_\nu^2 \to \mathrm{var}_P(z_t)$ as $\nu \nearrow$

Some intuition

- Simplified estimator: $\theta_0 = \mathbb{E}_P(z_t)$

$$\hat{\mu}_n(\nu) = \frac{1}{n}\sum_{t=1}^{n} \frac{z_t}{1 + |z_t|^2/\nu}$$

- Finite-Sample: a) $\sup_z |z|/(1 + |z|^2/\nu) \leq \sqrt{\nu}/2$, b) Bernstein's inequality:

$$\mathbb{P}\left(\sup_{z_t \in \mathcal{O}_n, t > n_P} |\hat{\mu}_n(\nu) - \mu(\nu)| \geq \frac{\sqrt{\nu} n_o}{n} + \frac{n_P}{n}\frac{x}{\sqrt{n_P}}\right) \leq 2\exp\left(-\frac{x^2}{2}\sigma_\nu^2 + 2/3\sqrt{\nu/n_P}x\right)$$

when $z_t \sim P$, $1 \leq t \leq n_P$, are iid. $\sigma_\nu^2 \to \mathrm{var}_P(z_t)$ as $\nu \nearrow$

- Two biases:
  - outlier: $\sqrt{\nu} n_o / n$
  - asymmetry: $\mu(\nu) - \theta_0 = \mathbb{E}_P[\hat{\mu}_n(\nu)] - \theta_0$

Some intuition

- Asymmetry bias is at most $\mathbb{E}_P(|z_t|^3)/\nu$:

$$\mu(\nu) = \theta_0 - \frac{1}{\nu}\mathbb{E}_P\left(\frac{z_t^3}{1 + z_t^2/\nu}\right)$$

Some intuition

- Asymmetry bias is at most $\mathbb{E}_P(|z_t|^3)/\nu$:

$$\mu(\nu) = \theta_0 - \frac{1}{\nu}\mathbb{E}_P\left(\frac{z_t^3}{1 + z_t^2/\nu}\right)$$

- Bias correction $\tilde{\mu}(\nu) = 2\mu(\nu) - \mu(\nu/2)$:

$$\tilde{\mu}(\nu) = \theta_0 - \frac{1}{\nu^2}\mathbb{E}_P\left(\frac{z_t^5}{(1 + z_t^2/\nu)(1 + 2z_t^2/\nu)}\right)$$

Some intuition

- Asymmetry bias is at most $\mathbb{E}_P(|z_t|^3)/\nu$:

$$\mu(\nu) = \theta_0 - \frac{1}{\nu}\mathbb{E}_P\left(\frac{z_t^3}{1 + z_t^2/\nu}\right)$$

- Bias correction $\tilde{\mu}(\nu) = 2\mu(\nu) - \mu(\nu/2)$:

$$\tilde{\mu}(\nu) = \theta_0 - \frac{1}{\nu^2}\mathbb{E}_P\left(\frac{z_t^5}{(1 + z_t^2/\nu)(1 + 2z_t^2/\nu)}\right)$$

- Repeat bias correction $\tilde{\tilde{\mu}}(\nu) = 2\tilde{\mu}(\nu) - \tilde{\mu}(\nu/2)$:

$$\tilde{\mu}(\nu) = \theta_0 + \frac{2}{\nu^2}\mathbb{E}_P\left(\frac{z_t^5(1 - z_t^4/\nu)}{(1 + z_t^2/\nu)(1 + 2z_t^2/\nu)(1 + 2z_t^2/\nu)(1 + 4z_t^2/\nu)}\right)$$

numerator has 3 roots, better small sample properties (simulations)

Trading-off biases

- Two biases:
  - Outlier bias: $\sqrt{\nu} n_o / n$
  - Asymmetry bias: $1/\nu^2$

Trading-off biases

- Two biases:
  - Outlier bias: $\sqrt{\nu} n_o / n$
  - Asymmetry bias: $1/\nu^2$
- Optimal choice $\nu \asymp (n/n_o)^{2/5}$

Trading-off biases

- Two biases:
  - Outlier bias: $\sqrt{\nu} n_o / n$
  - Asymmetry bias: $1/\nu^2$
- Optimal choice $\nu \asymp (n/n_o)^{2/5}$
- Asymptotic normality requires: $\sqrt{n}/\nu^2 = o(1), \sqrt{\nu/n} n_o = o(1) \Rightarrow n_o = o(n^{3/8})$
  - n_o cannot increase too quickly...
  - P symmetric, need: $n_o = o(n^{1/2})$
  - $\nu \asymp n^{1/4} \log(n)$ implies $n_o = o(n^{3/8}/\sqrt{\log(n)})$

Trading-off biases

- Two biases:
  - Outlier bias: $\sqrt{\nu} n_o / n$
  - Asymmetry bias: $1/\nu^2$
- Optimal choice $\nu \asymp (n/n_o)^{2/5}$
- Asymptotic normality requires: $\sqrt{n}/\nu^2 = o(1), \sqrt{\nu/n} n_o = o(1) \Rightarrow n_o = o(n^{3/8})$
  - n_o cannot increase too quickly...
  - P symmetric, need: $n_o = o(n^{1/2})$
  - $\nu \asymp n^{1/4} \log(n)$ implies $n_o = o(n^{3/8}/\sqrt{\log(n)})$
- Undersmoothing (no bias correction): $n_o = o(n^{1/4})$

Finite Sample Results

$$Q_\nu(\psi;\theta) = \mathbb{E}_P\left[(\nu + p)\log\left(1 + \|g(z_t;\theta) - \mu\|^2_{\Sigma^{-1}}/\nu\right)\right] + \log|\Sigma| + \frac{\kappa_1}{\nu}\|\mu\|^2_{\Sigma^{-1}} + \frac{\kappa_2}{\nu}\mathrm{trace}(\Sigma).$$

- Q_ν: population analog of Q_n with $n_o = 0$, let $\psi(\theta;\nu)$ minimize Q_ν

Finite Sample Results

$$Q_\nu(\psi;\theta) = \mathbb{E}_P\left[(\nu+p)\log\left(1+\|g(z_t;\theta)-\mu\|^2_{\Sigma^{-1}}/\nu\right)\right] + \log|\Sigma| + \frac{\kappa_1}{\nu}\|\mu\|^2_{\Sigma^{-1}} + \frac{\kappa_2}{\nu}\mathrm{trace}(\Sigma).$$

- Q_ν: population analog of Q_n with $n_o = 0$, let $\psi(\theta;\nu)$ minimize Q_ν
- If i. $\mathbb{E}_P[\|g(z_t;\theta)\|^2] < \infty$, $\forall\theta$,
  ii. $\|g(z_t;\theta_1) - g(z_t;\theta_2)\| \leq G_t\|\theta_1-\theta_2\|$, $\mathbb{E}_P(\|G_t\|^2) < \infty$, then, for iid data:

$$\mathbb{P}\left(\sup_{\theta\in\Theta}\sup_{z_t\in\mathcal{O}_n, t>n_P}\left\{Q_\nu(\hat{\psi}_n(\theta;\nu);\theta) - Q_\nu(\psi(\theta;\nu);\theta)\right\} \geq C_{\mathcal{O}}\frac{n_o(\nu+p)}{n}[1+\log(n)]\right.$$
$$\left. + L\frac{n_P}{n}(\nu+p)\log(1+\nu p)\left[\sqrt{\frac{x}{n_P}} + \frac{x}{n_P} + \sqrt{\frac{C_n}{n_P}} + \frac{C_n}{n_P}\right]\right) \leq 4\exp(-x),$$

where $C_n = 1 + (k+2p^2)[\log(p) + \log(\nu) + \log(n_P)]$

Large Sample Results: Oracle Equivalence

- Further assume that $\sup_{\theta \in \Theta} \mathbb{E}_P[\|g(z_t;\theta)\|^4] < \infty$, and

$$n_o = o\left(\frac{n}{\nu \log(n)}\right), \; \nu \log(\nu) = o\left(\sqrt{\frac{n}{\log(n)}}\right).$$

- Let $M_{r,\delta} = \max\left(\mathbb{E}_P(\|g(z_t;\theta_0)\|^{r+\delta}), \mathbb{E}_P(\|G_t\|^{r+\delta})\right)$ and $\overline{g}_{n_P}(\theta) = \frac{1}{n_P}\sum_{t=1}^{n_P} g(z_t;\theta)$

Large Sample Results: Oracle Equivalence

- Further assume that $\sup_{\theta \in \Theta} \mathbb{E}_P[\|g(z_t;\theta)\|^4] < \infty$, and

$$n_o = o\left(\frac{n}{\nu \log(n)}\right), \ \nu \log(\nu) = o\left(\sqrt{\frac{n}{\log(n)}}\right).$$

- Let $M_{r,\delta} = \max\left(\mathbb{E}_P(\|g(z_t;\theta_0)\|^{r+\delta}), \mathbb{E}_P(\|G_t\|^{r+\delta})\right)$ and $\overline{g}_{n_P}(\theta) = \frac{1}{n_P}\sum_{t=1}^{n_P} g(z_t;\theta)$
- If $M_{3,\delta} < \infty$ for some $\delta > 0$:

$$\sup_{\theta \in \Theta}\left(\sup_{z_t \in \mathcal{O}_n, t > n_P} \|\hat{\mu}_n(\theta;\nu) - \overline{g}_{n_P}(\theta)\|\right) = O_p\left(\max\left[\frac{1}{\nu}, \frac{\sqrt{\nu} n_o}{n}\right]\right)$$

Large Sample Results: Oracle Equivalence

- Further assume that $\sup_{\theta\in\Theta} \mathbb{E}_P[\|g(z_t;\theta)\|^4] < \infty$, and

$$n_o = o\left(\frac{n}{\nu \log(n)}\right), \; \nu \log(\nu) = o\left(\sqrt{\frac{n}{\log(n)}}\right).$$

- Let $M_{r,\delta} = \max\left(\mathbb{E}_P(\|g(z_t;\theta_0)\|^{r+\delta}), \mathbb{E}_P(\|G_t\|^{r+\delta})\right)$ and $\bar{g}_{n_P}(\theta) = \frac{1}{n_P}\sum_{t=1}^{n_P} g(z_t;\theta)$
- If $M_{3,\delta} < \infty$ for some $\delta > 0$:

$$\sup_{\theta\in\Theta}\left(\sup_{z_t\in\mathcal{O}_n, t>n_P} \|\hat{\mu}_n(\theta;\nu) - \bar{g}_{n_P}(\theta)\|\right) = O_p\left(\max\left[\frac{1}{\nu}, \frac{\sqrt{\nu}n_o}{n}\right]\right)$$

- If $M_{5,\delta} < \infty$ for some $\delta > 0$:

$$\sup_{\theta\in\Theta}\left(\sup_{z_t\in\mathcal{O}_n, t>n_P} \|\tilde{\mu}_n(\theta;\nu) - \bar{g}_{n_P}(\theta)\|\right) = O_p\left(\max\left[\frac{1}{\nu^2}, \frac{\sqrt{\nu}n_o}{n}\right]\right).$$

Large Samples: Oracle Equivalence (estimator)

- Assume standard regularity conditions for $1 \leq t \leq n_P$, and

$$\frac{\sqrt{n}}{\nu^2} = o(1), \text{ and } \sqrt{\frac{\nu}{n}} n_o = o(1).$$

Large Samples: Oracle Equivalence (estimator)

- Assume standard regularity conditions for $1 \leq t \leq n_P$, and

$$\frac{\sqrt{n}}{\nu^2} = o(1), \text{ and } \sqrt{\frac{\nu}{n}} n_o = o(1).$$

- Let $\hat{\theta}_{n_P}$ minimize $\|\overline{g}_{n_P}(\theta)\|_{W_n}$; $\tilde{\theta}_n$ minimize $\|\tilde{\mu}_n(\theta;\nu)\|_{W_n}$ then:

$$\sup_{z_t \in \mathcal{O}_n, t > n_P} \|\sqrt{n_P}(\tilde{\theta}_n - \hat{\theta}_{n_P})\| = o_p(1), \quad \text{and} \quad \sqrt{n_P}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$$

for any sequence $(z_t)_{t > n_P} \in \mathcal{O}_n$, $V = \text{avar}(\hat{\theta}_{n_P})$

Monte Carlo

- Simulation design – for $1 \leq t \leq n_P$

$$y_t = \theta_0 + \theta_1 x_{1t} + \theta_2 x_{2t} + \theta_3 x_{3t} + e_t,$$

- $n = 150$, $x_{1t}, x_{2t}, x_{3t}, e_t = (\chi_5^2 - 5)/\sqrt{10}$. $\theta_0 = (0, 1, 1, 1)$

Monte Carlo

- Simulation design – for $1 \leq t \leq n_P$

$$y_t = \theta_0 + \theta_1 x_{1t} + \theta_2 x_{2t} + \theta_3 x_{3t} + e_t,$$

- $n = 150$, $x_{1t}, x_{2t}, x_{3t}, e_t = (\chi^2_5 - 5)/\sqrt{10}$. $\theta_0 = (0, 1, 1, 1)$
- Outliers: $y_t = x_t' \theta_\dagger$, $\theta_\dagger = (0, 1/2, 1/2, 1/2)$. $x_{1t} = x_{2t} = x_{3t} = \sqrt{n}$
- $n_o = 0, 1, 5, 10, \ldots$

Monte Carlo

- Simulation design – for $1 \leq t \leq n_P$

$$y_t = \theta_0 + \theta_1 x_{1t} + \theta_2 x_{2t} + \theta_3 x_{3t} + e_t,$$

- $n = 150$, $x_{1t}, x_{2t}, x_{3t}, e_t = (\chi^2_5 - 5)/\sqrt{10}$. $\theta_0 = (0, 1, 1, 1)$
- Outliers: $y_t = x_t'\theta_\dagger$, $\theta_\dagger = (0, 1/2, 1/2, 1/2)$. $x_{1t} = x_{2t} = x_{3t} = \sqrt{n}$
- $n_o = 0, 1, 5, 10, \ldots$
- Several estimators:
    - OLS,
    - oracle estimator,
    - Robust M-estimator with Huber loss,
    - This paper's: $\hat{\theta}_n$, $\tilde{\theta}_n$, $\tilde{\tilde{\theta}}_n$
    - Undersmoothing: no bias correction, ν^2

Monte Carlo: $n = 150$, $n_o = 0$

	100 × RMSE							Rejection Rate						
	$n_o = 0$													
	$\hat{\theta}_n^{ols}$	$\hat{\theta}_{nP}^{ols}$	$\hat{\theta}_n^{rlm}$	$\hat{\theta}_n$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n$	$\hat{\theta}_n^{un}$	$\hat{\theta}_n^{ols}$	$\hat{\theta}_{nP}^{ols}$	$\hat{\theta}_n^{rlm}$	$\hat{\theta}_n$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n$	$\hat{\theta}_n^{un}$
θ_0	8.05	8.05	12.00	11.84	9.31	8.11	7.94	0.04	0.04	0.24	0.29	0.14	0.05	0.06
θ_1	8.00	8.00	7.15	7.97	7.79	7.78	7.92	0.06	0.06	0.06	0.11	0.08	0.07	0.06
θ_2	8.10	8.10	7.46	8.45	8.21	8.11	8.06	0.04	0.04	0.05	0.10	0.06	0.05	0.05
θ_3	8.19	8.19	7.43	8.55	8.30	8.16	8.14	0.06	0.06	0.06	0.10	0.07	0.06	0.06

Average $\hat{\nu}_n$: 35.85, 16.00, 11.00, 10.71 for $n_0 = 0, 1, 5, 10$ respectively. Each $\hat{\nu}_n$ is selected on a grid $[\nu_0, \ldots, \nu_J]$ where $\nu_0 = 8.77$, $\nu_J = 584.69$.

Monte Carlo: $n = 150$, $n_o = 1$

	100 × RMSE							Rejection Rate						
	$n_o = 1$													
	$\hat{\theta}_n^{ols}$	$\hat{\theta}_{n_P}^{ols}$	$\hat{\theta}_n^{rlm}$	$\hat{\theta}_n$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n$	$\hat{\theta}_n^{un}$	$\hat{\theta}_n^{ols}$	$\hat{\theta}_{n_P}^{ols}$	$\hat{\theta}_n^{rlm}$	$\hat{\theta}_n$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n$	$\hat{\theta}_n^{un}$
θ_0	10.71	8.04	13.01	14.18	10.97	8.52	10.32	0.03	0.04	0.20	0.46	0.23	0.08	0.08
θ_1	38.57	8.07	15.23	8.27	7.97	7.87	32.24	0.00	0.06	0.01	0.14	0.10	0.07	0.40
θ_2	38.39	8.11	15.09	8.73	8.36	8.14	32.08	0.01	0.04	0.01	0.12	0.06	0.06	0.38
θ_3	39.94	8.20	15.75	8.83	8.49	8.27	33.47	0.00	0.06	0.00	0.12	0.09	0.07	0.39

Average $\hat{\nu}_n$: 35.85, 16.00, 11.00, 10.71 for $n_0 = 0, 1, 5, 10$ respectively. Each $\hat{\nu}_n$ is selected on a grid $[\nu_0, \ldots, \nu_J]$ where $\nu_0 = 8.77$, $\nu_J = 584.69$.

Monte Carlo: $n = 150$, $n_o = 5$

	100 × RMSE							Rejection Rate						
	$n_o = 5$													
	$\hat{\theta}_n^{ols}$	$\hat{\theta}_{nP}^{ols}$	$\hat{\theta}_n^{rlm}$	$\hat{\theta}_n$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n$	$\hat{\theta}_n^{un}$	$\hat{\theta}_n^{ols}$	$\hat{\theta}_{nP}^{ols}$	$\hat{\theta}_n^{rlm}$	$\hat{\theta}_n$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n$	$\hat{\theta}_n^{un}$
θ_0	11.98	8.14	16.57	16.98	13.38	9.82	13.45	0.10	0.04	0.24	0.59	0.38	0.13	0.16
θ_1	47.57	8.40	47.17	9.02	8.62	8.40	46.72	0.99	0.06	0.99	0.12	0.08	0.06	0.99
θ_2	47.48	8.26	48.25	9.28	8.80	8.53	47.14	0.99	0.04	1.00	0.12	0.05	0.03	1.00
θ_3	49.17	8.28	49.48	9.33	8.94	8.72	48.65	0.98	0.06	0.98	0.10	0.08	0.04	0.98

Average $\hat{\nu}_n$: 35.85, 16.00, 11.00, 10.71 for $n_0 = 0, 1, 5, 10$ respectively. Each $\hat{\nu}_n$ is selected on a grid $[\nu_0, \ldots, \nu_J]$ where $\nu_0 = 8.77$, $\nu_J = 584.69$.

Monte Carlo: $n = 150$, $n_o = 10$

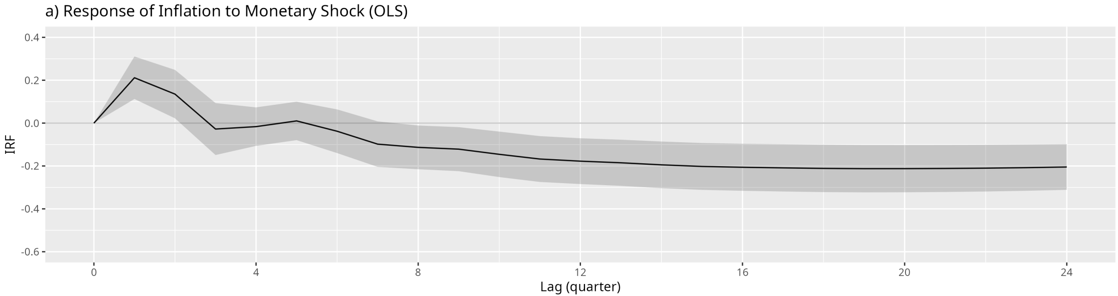
	100 × RMSE							Rejection Rate						
	$n_o = 10$													
	$\hat{\theta}_n^{ols}$	$\hat{\theta}_{nP}^{ols}$	$\hat{\theta}_n^{rlm}$	$\hat{\theta}_n$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n$	$\hat{\theta}_n^{un}$	$\hat{\theta}_n^{ols}$	$\hat{\theta}_{nP}^{ols}$	$\hat{\theta}_n^{rlm}$	$\hat{\theta}_n$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n$	$\hat{\theta}_n^{un}$
θ_0	12.21	8.21	17.33	16.78	13.27	10.35	14.13	0.09	0.04	0.23	0.47	0.22	0.07	0.17
θ_1	49.14	8.54	48.38	10.22	11.68	19.76	48.65	0.99	0.04	0.99	0.01	0.01	0.09	1.00
θ_2	49.05	8.31	49.67	10.76	12.40	20.28	48.92	0.99	0.04	0.99	0.01	0.01	0.09	1.00
θ_3	50.52	8.51	50.70	11.04	13.00	20.96	50.19	0.98	0.06	0.98	0.00	0.01	0.09	0.99

Average $\hat{\nu}_n$: 35.85, 16.00, 11.00, 10.71 for $n_0 = 0, 1, 5, 10$ respectively. Each $\hat{\nu}_n$ is selected on a grid $[\nu_0, \ldots, \nu_J]$ where $\nu_0 = 8.77$, $\nu_J = 584.69$.

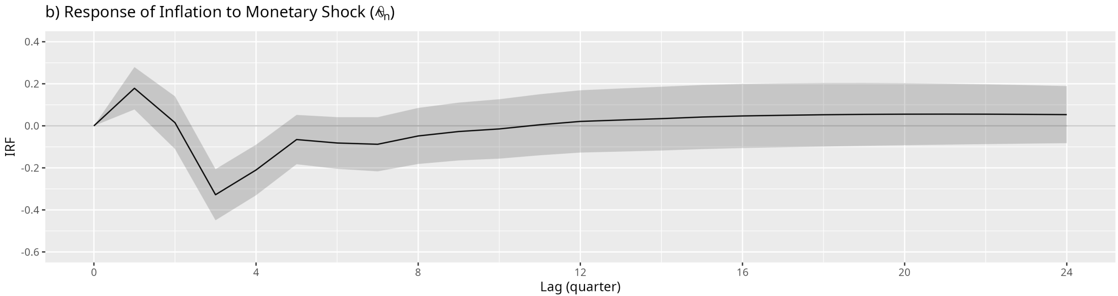
Back to the Price Puzzle

- Re-estimate the model using OLS, $\hat{\theta}_n$, $\tilde{\theta}_n$, $\tilde{\tilde{\theta}}_n$
- Same data
- Same model

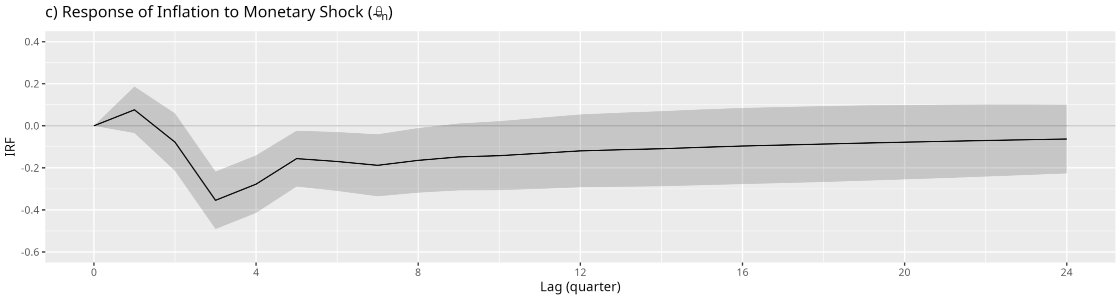
Price Puzzle: OLS



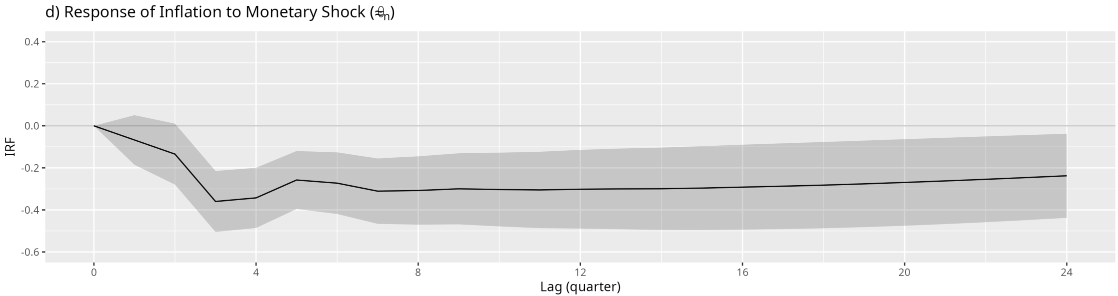
Price Puzzle: Robust, No Bias Correction



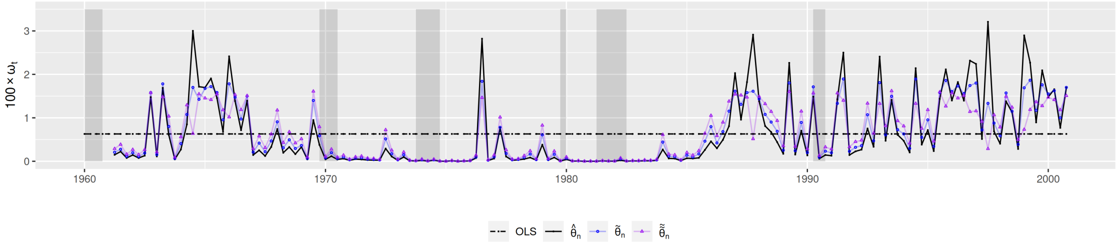
Price Puzzle: Robust, Bias Correction



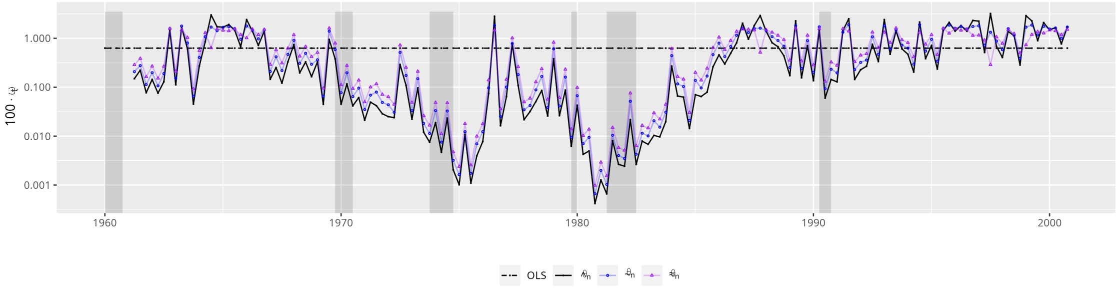
Price Puzzle: Robust, Repeat Bias Correction



Price Puzzle: Estimation Weights



Price Puzzle: Estimation Weights (log scale)



Conclusion

- Misspecification can occur
  - don't always know which t are involved
  - diagnostics useful but not definitive
- Robust estimation:
  - more resilient
  - potentially biased/inconsistent
- This paper: simple estimates, bias correction
  - some finite and large sample results
  - robustness to leveraged outliers
  - weights: make the results transparent

References I

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1):387–422.

Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, 96(455):1022–1030.

Carroll, R. J. and Welsh, A. H. (1988). A note on asymmetry and robustness in linear regression. *The American Statistician*, 42(4):285–287.

Castelnuovo, E. and Surico, P. (2010). Monetary policy, inflation expectations and the price puzzle. *The Economic Journal*, 120(549):1262–1283.

Eichenbaum, M. (1992). Comment on 'interpreting the macroeconomic time series facts: The effects of monetary policy': by christopher sims. *European Economic Review*, 36(5):1001–1011.

Hamilton, L. C. (1992). How robust is robust regression? *Stata Technical Bulletin*, 1(2).

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101.

Jaeckel, L. A. (1971). Robust estimates of location: Symmetry and asymmetric contamination. *The Annals of Mathematical Statistics*, 42(3):1020–1034.

References II

Laforgue, P., Staerman, G., and Clémençon, S. (2021). Generalization bounds in the presence of outliers: a median-of-means study. In *International Conference on Machine Learning*, pages 5937–5947. PMLR.

Lecué, G. and Lerasle, M. (2020). Robust machine learning by median-of-means: Theory and practice. *The Annals of Statistics*, 48(2):906 – 931.

Newey, W. K. and Steigerwald, D. G. (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica: Journal of the Econometric Society*, pages 587–599.

Rusnák, M., Havranek, T., and Horváth, R. (2013). How to solve the price puzzle? a meta-analysis. *Journal of Money, Credit and Banking*, 45(1):37–70.

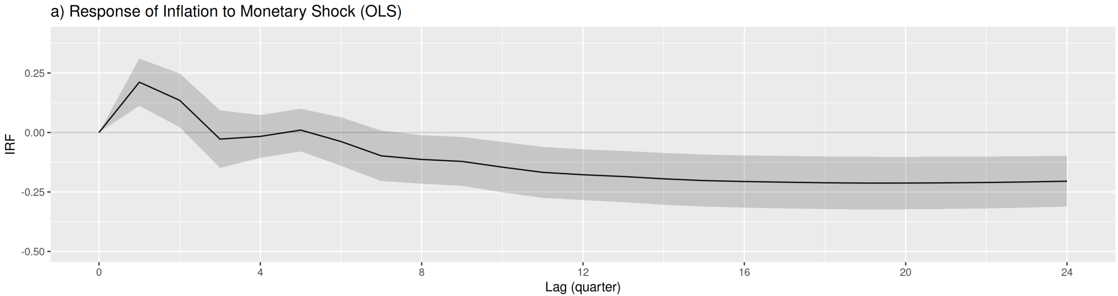
Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36(5):975–1000.

Stock, J. H. and Watson, M. W. (2001). Vector autoregressions. *Journal of Economic perspectives*, 15(4):101–115.

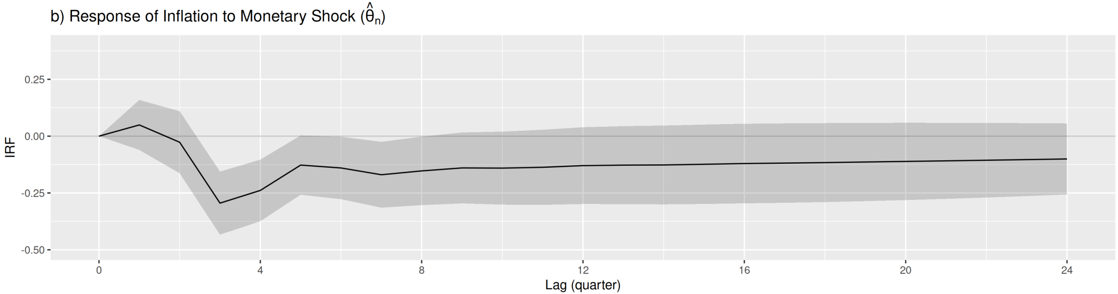
Back to the Price Puzzle

- Re-estimate with a larger $\nu = 15$ vs. $\nu = 8.99$ in the main results

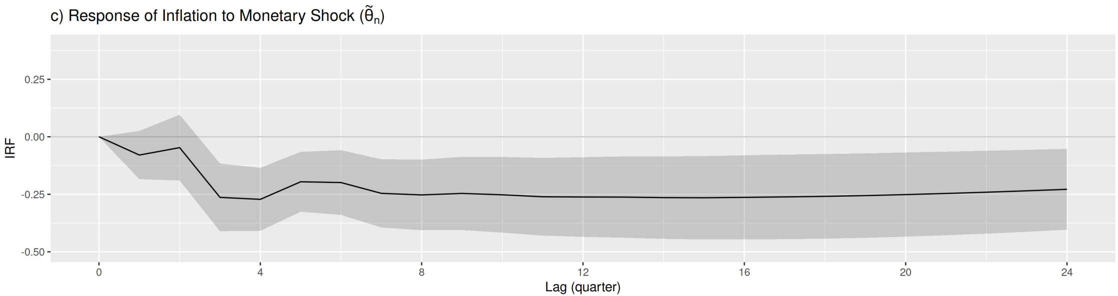
Price Puzzle: OLS



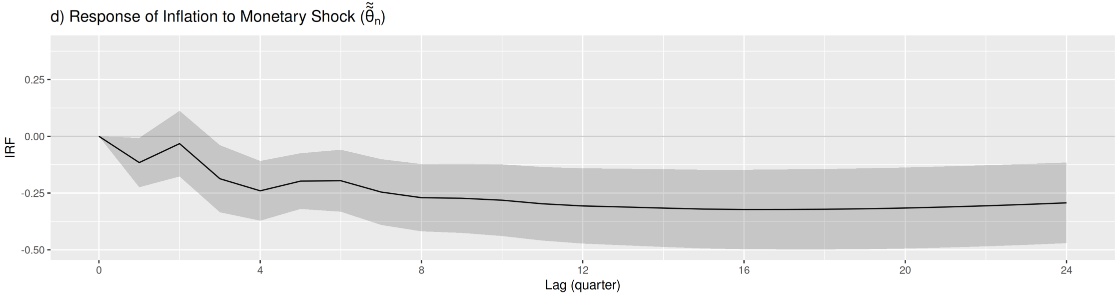
Price Puzzle: Robust, No Bias Correction



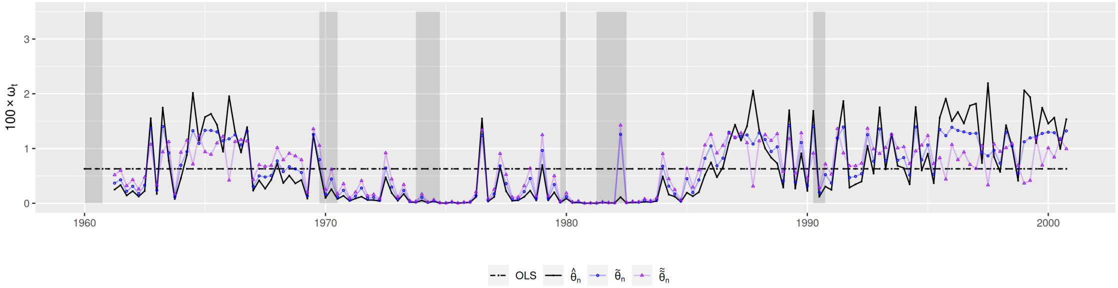
Price Puzzle: Robust, Bias Correction



Price Puzzle: Robust, Repeat Bias Correction



Price Puzzle: Estimation Weights



Price Puzzle: Estimation Weights (log scale)