When Instruments Break: Choosing the Optimal Estimation Fraction Under a Change in Exogeneity

Francis J. DiTraglia

University of Pennsylvania

April 7, 2014

Preface

1. Joint work with Otilia Boldea, Tilburg University
2. Preliminary: no paper yet, this is the first presentation.
3. Comments greatly appreciated!

What to do if your instruments *break* at an unknown date?

Linear Time-series Regression

$y_t = x_t'\theta^o + u_t$ where x_t is a $p \times 1$ vector of endogenous explanatory variables.

Instrumental Variables

z_t is a $q \times 1$ vector of relevant instruments for x_t

Break in Endogeneity

For all time periods through T^o the instruments z_t are exogenous. *After* T^o, they *may be endogenous*. We don't know T^o.

We're *not* going to do a break test.

A Possible Procedure

1. Use a structural break test to locate T^o.
2. Estimate θ using data through estimated break date $\widehat{T}$.

Problems With This Approach

1. It's indirect: break tests aim to find T^o, *not* estimate θ.
2. If T^o were known, we *still* might want to use data *beyond this point*: there's a bias-variance tradeoff.

Our Proposal: Minimum Risk Estimation Fraction

Basic Idea

Choose the estimation fraction that yields the lowest weighted mean-squared error (MSE) estimator of θ.

Intuition

Risk-optimal break fraction should depend on *how endogenous* z_t becomes after the break, sample size, and various other quantities.

Asymptotic Framework

Use AMSE to approximate finite-sample MSE, and local asymptotics to get a limiting bias-variance tradeoff.

Asymptotic Framework

Local Endogeneity

$$E[z_t u_t] = \begin{cases} 0 & t \leq T^o \\ c/\sqrt{T} & t > T^o \end{cases}$$

AMSE for Vector $\widetilde{\theta}$

$$\text{AMSE}(\widetilde{\theta}) = \lim_{T \to \infty} E\left[T \left(\widetilde{\theta} - \theta^o \right)' \Sigma \left(\widetilde{\theta} - \theta^o \right) \right]$$

(Σ is a weighting matrix)

Some Notation

Estimation Fraction λ

Sample is split in two: $t = 1, \ldots, \lfloor T\lambda \rfloor$ and $t = T - \lfloor T\lambda \rfloor, \ldots, T$

Subscripts

1λ = first part, 2λ = second part

Asymptotics for GMM Estimator

Limit Distribution of Estimator

$$\sqrt{T}\left[\widehat{\theta}_{1\lambda}(W)-\theta^{o}\right]\Rightarrow (QWQ')^{-1}QW\left[\frac{V^{1/2}B_q(\lambda)}{\lambda}+\mathbf{1}\left\{\lambda>\lambda^{o}\right\}\left(\frac{\lambda-\lambda^{o}}{\lambda}\right)c\right].$$

Optimal Weighting Matrix

$$W=\text{AVAR}\left[\frac{Z_{1\lambda}'U_{1\lambda}}{T_{1\lambda}}\cdot\sqrt{T}\right]=\left(\frac{V^{1/2}}{\lambda}\right)Var\left(B\left(\lambda\right)\right)\left(\frac{V^{1/2}}{\lambda}\right)'=\frac{\lambda V}{\lambda^{2}}=\frac{V}{\lambda},$$

Weighting Matrices

GMM Estimation

We use the optimal weighting matrix.

AMSE Definition

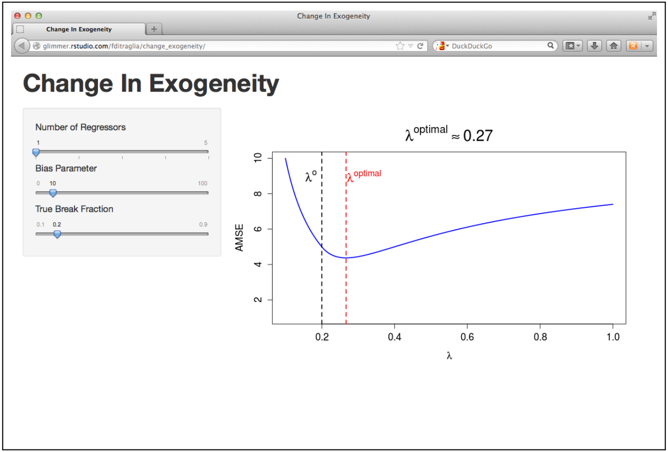
We use the “natural” choice: $\Sigma = QV^{-1}Q'$

Theorem (Asymptotic Mean-squared Error)

$$AMSE(\lambda) = \frac{p}{\lambda} + \mathbf{1}\{\lambda > \lambda^o\}\left(\frac{\lambda - \lambda^o}{\lambda}\right)^2 b$$

$$\begin{aligned}
\lambda &= \textit{Estimation Fraction} \\
\lambda^o &= \textit{True Pre-Break Fraction} \\
b &= \textit{Squared Bias Parameter} \quad \left(c' V^{-1/2} P_S V^{-1/2} c\right) \\
p &= \textit{Number of Regressors}
\end{aligned}$$

http://glimmer.rstudio.com/fditraglia/change_exogeneity/



Theorem (Optimal Estimation Fraction)

Suppose that $\lambda^o \in (0,1)$. Then, regardless of the size of the bias parameter, b, and the number of regressors, p, the estimation fraction λ^* that minimizes AMSE is always strictly greater than λ^o, the true break fraction. Specifically...

Theorem (Optimal Estimation Fraction)

$$\lambda^* = \begin{cases} \lambda^o \left[1 - \dfrac{p}{2b\lambda^o}\right]^{-1} & b \geq b_{min} \\ 1 & b < b_{min} \end{cases}$$

$$b_{min} = \frac{p}{2\lambda^o (1 - \lambda^o)}$$

$$\begin{aligned} \lambda^* &= \textit{AMSE-Optimal Estimation Fraction} \\ \lambda^o &= \textit{True Pre-Break Fraction} \\ b &= \textit{Squared Bias Parameter} \quad \left(c' V^{-1/2} P_S V^{-1/2} c\right) \\ p &= \textit{Number of Regressors} \end{aligned}$$

Simulation Experiment

Do Local Asymptotics "Work?"

Compare AMSE formulas at *true* DGP parameters to (scaled) *exact* finite sample MSE calculated by simulation.

Important!

This experiment uses an *infeasible* criterion: choosing λ when b and λ^o are *known*. This is a "sanity check" for the asymptotics.

DGP for Simulation Experiment

Fix $T = 200$ and vary γ and T^o

$$
\begin{aligned}
y_t &= x_t + u_t \\
x_t &= 0.5 z_t + v_t
\end{aligned}
\qquad
\begin{aligned}
z_t &= 0.2 z_{t-1} + \epsilon_t^{(z)} \\
u_t &= \epsilon_t^{(u)} \\
v_t &= 0.2 v_{t-1} + \epsilon_t^{(v)}
\end{aligned}
$$

$$\left[\epsilon_t^{(z)}, \epsilon_t^{(u)}, \epsilon_t^{(v)}\right]' \sim \text{inid } \mathrm{N}(0, \Omega_t)$$

$$\Omega_t = \begin{bmatrix} 1 & \gamma \mathbf{1}\{t > T^o\} & 0 \\ \gamma \mathbf{1}\{t > T^o\} & 1 & 0.2 \\ 0 & 0.2 & 1 \end{bmatrix}$$

Simulation Specifics

Goal of the Exercise

Estimate the coefficient θ on x_t with minimum MSE.

All Possible Estimators

Sequence of IV estimators $\widetilde{\theta}_\tau = \left(\sum_{t=1}^{\tau} z_t y_t\right) / \left(\sum_{t=1}^{\tau} z_t x_t\right)$ for all $\tau = 1, 2, \ldots, T$.

Comparison

When properly re-scaled, does the MSE calculated from the simulation agree with our theoretical AMSE formula?

Specializing the Asymptotic Results to This Example

$$AMSE(\lambda) = \frac{1}{\lambda} + \mathbf{1}\left\{\lambda > \lambda^o\right\}\left(\frac{\lambda - \lambda^o}{\lambda}\right)^2 T\gamma^2$$

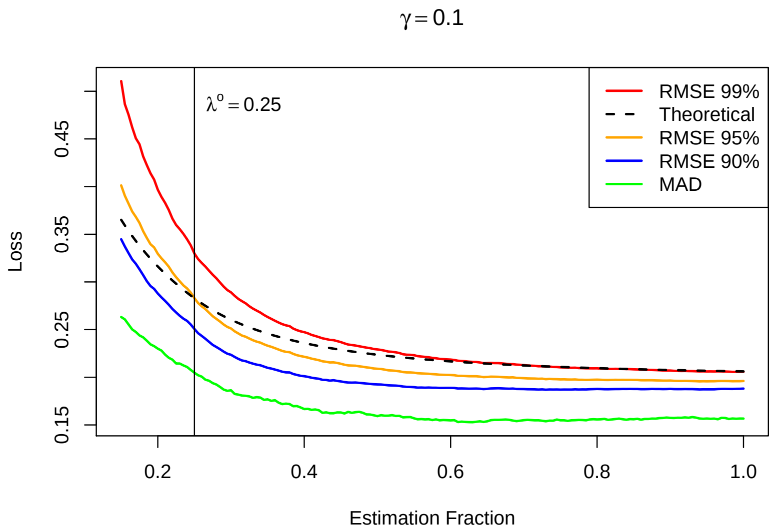
- Single regressor: $p = 1$
- $\gamma = Cov\left(z_t, u_t \,|t > T^o\right)$ plays the role of $c/\sqrt{T}$
- All errors have unit variance $\implies$ $b = T\gamma^2$

Re-scaling Finite Sample MSE

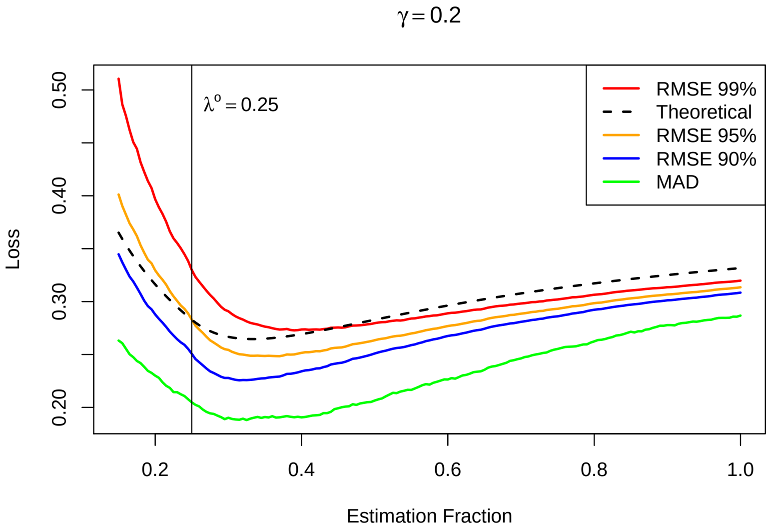
In this example, Σ is a scalar and $T = 200$

$$
\begin{aligned}
AMSE &= \lim_{T\to\infty} E\left[T \left(\widetilde{\theta} - \theta^o\right)' \Sigma \left(\widetilde{\theta} - \theta^o\right)\right] \\
&= \lim_{T\to\infty} T \left[\frac{0.5^2}{1-0.2^2}\right] E\left[\left(\widetilde{\theta} - \theta^o\right)^2\right] \\
&\approx 52 \cdot \mathrm{MSE}(\widetilde{\theta})
\end{aligned}
$$

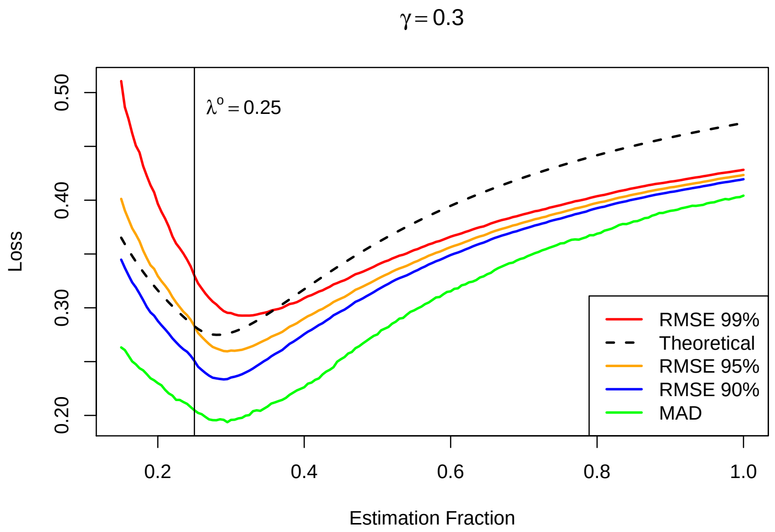
```r
set.seed(4637)
loss_plot_cpp(N_reps = 10000, g = 0.1, T0 = 50, T1 = 150)
```



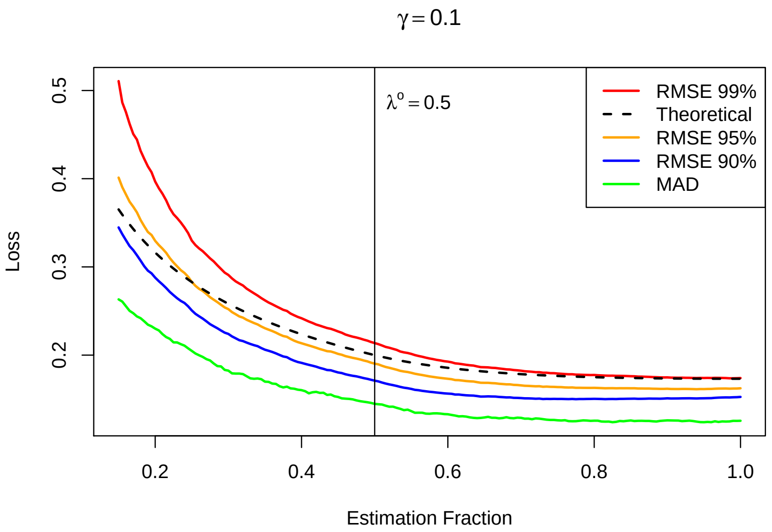
```r
set.seed(4637)
loss_plot_cpp(N_reps = 10000, g = 0.2, T0 = 50, T1 = 150)
```



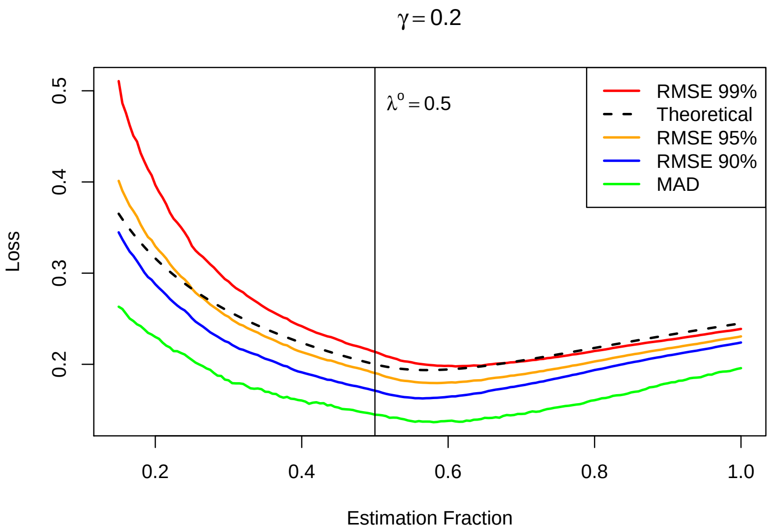
```r
set.seed(4637)
loss_plot_cpp(N_reps = 10000, g = 0.3, T0 = 50, T1 = 150, legend.pos =
```



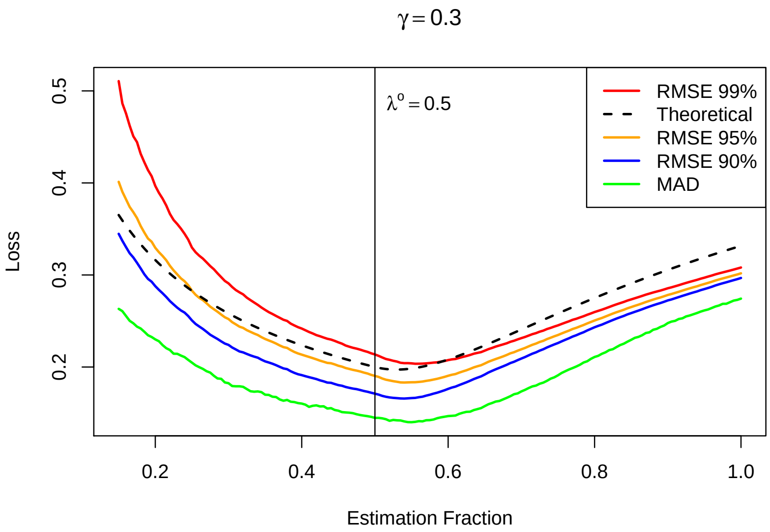
```r
set.seed(4637)
loss_plot_cpp(N_reps = 10000, g = 0.1, T0 = 100, T1 = 100)
```



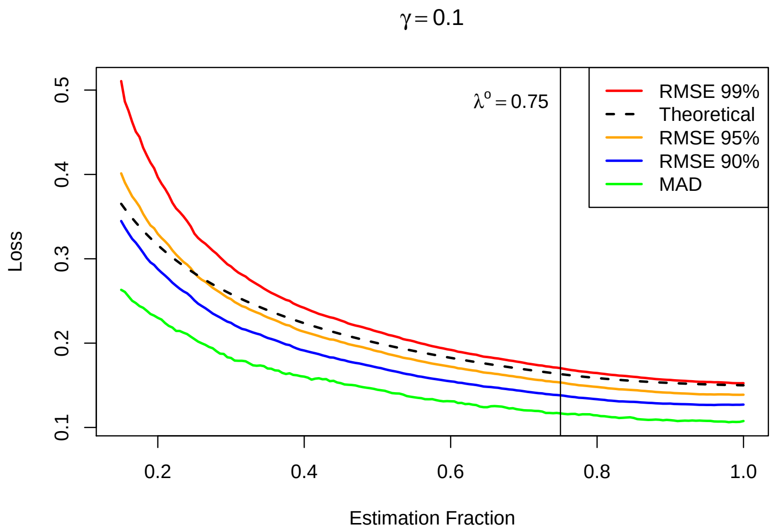
```r
set.seed(4637)
loss_plot_cpp(N_reps = 10000, g = 0.2, T0 = 100, T1 = 100)
```



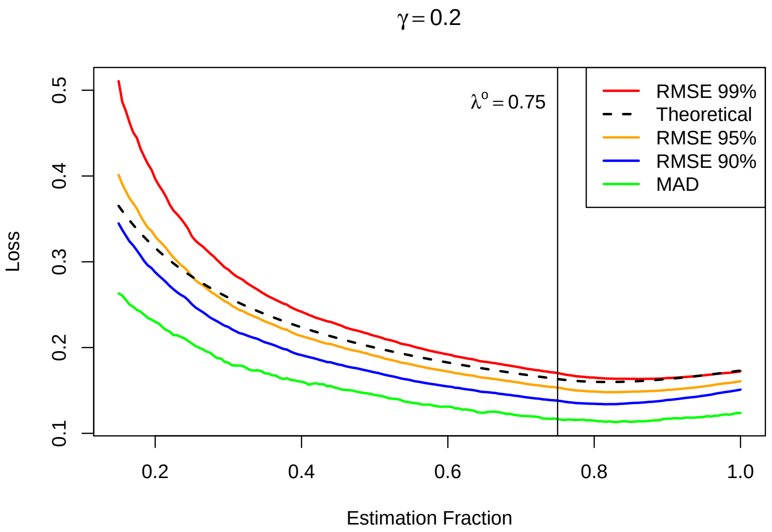
```r
set.seed(4637)
loss_plot_cpp(N_reps = 10000, g = 0.3, T0 = 100, T1 = 100)
```



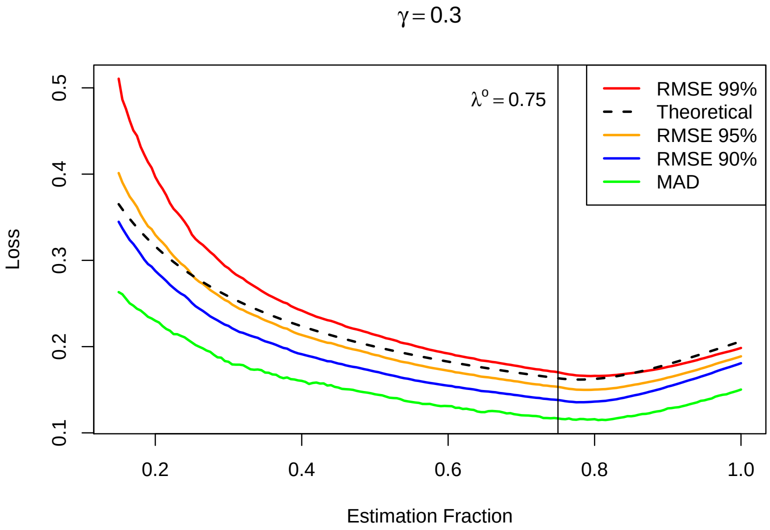
```r
set.seed(4637)
loss_plot_cpp(N_reps = 10000, g = 0.1, T0 = 150, T1 = 50)
```



```r
set.seed(4637)
loss_plot_cpp(N_reps = 10000, g = 0.2, T0 = 150, T1 = 50)
```



```r
set.seed(4637)
loss_plot_cpp(N_reps = 10000, g = 0.3, T0 = 150, T1 = 50)
```



Estimating The Squared Bias Parameter

Interval Restriction

Suppose we know that $\lambda^o \in [\lambda^\ell, \lambda^u]$

Intuition

Plug estimator with estimation fraction λ^ℓ into *empirical moment conditions* for sample fraction $[\lambda^u, 1]$.

Asymptotically Unbiased!

$$\widehat{c}(\lambda^\ell, \lambda^u) \Rightarrow c + \Psi M(\lambda^\ell, \lambda^u)$$

What's Next: Feasible AMSE Estimator

Problem

AMSE formula depends on λ^o in addition to c.

Possible Solutions

1. "Least-favorable Alternative" – evaulate at λ^ℓ
2. Average over a prior for λ^o with support $[\lambda^\ell, \lambda^o]$

Use prior for λ^o to estimate c?

Each possible value for λ^o yields an associated plug in estimator of c. Should we average over these as well?

What's Next: Applications

We're still working on this, but have a few ideas from macroeconometrics: policy changes, changes in timing, changes in the information sets available to agents. Suggestions would be appreciated!