Narrative Restrictions and Proxies

Raffaella Giacomini (UCL and Federal Reserve Bank of Chicago)
Toru Kitagawa (Brown and UCL)
Matthew Read (UCL and Reserve Bank of Australia)

Motivation

- A leading approach to macroeconomic policy analysis: SVAR with "traditional" zero/sign restrictions on structural parameters
- Growing empirical literature uses new type of restrictions: "narrative restrictions - NR" = inequality restrictions on structural shocks on given dates

- Shock-sign: there was a positive monetary shock on given dates, e.g., “Volcker shock” (Antolin-Diaz & Rubio-Ramirez, 2018 - AR18).
- Historical decomposition: the change in interest rate on given dates was mostly due to a monetary shock (AR18)
- Magnitude: bounds on the size of shocks on given dates (Ludvigson, Ma & Ng, 2019 - LMN19)
- Shock-rank: the shock on a given date was the largest in the sample (Giacomini, Kitagawa & Read, 2021 - GKR21)

Empirical Examples

- Furlanetto and Robstad (2019, RED): the immigration shock was the most important contributor to the observed immigration in the mid 2000s in Norway by the opening up of the Norwegian labour market to Eastern European citizens.
- Laumer (2020, JEDC): NR on the sign of the government spending shock in periods corresponding to unanticipated increases in US military spending.
- Redl (2020, JIE): NR of positive macroeconomic uncertainty shocks around ’close’ general elections and for positive financial uncertainty shock during periods of ’financial distress’.

Challenges

- Nonstandard nature of NR both in identification and inference
- The current state of the art:
  - AR18 perform Bayesian inference
  - LMN19 use a bootstrap procedure, but (frequentist) validity is unknown
  - GKR21 analyzes identification and posterior sensitivity, proposes robust-Bayes inference, and shows it is asymptotically (frequentist) valid

Questions

- An alternative approach to exploit narratives is the proxy (external) instrumental variable (IV) approach (Stock 2008, Mertens & Ravn 2013, Stock & Watson 2018).
- E.g., transform shock-sign restrictions into a proxy IV; for $t = 1, \ldots, T$,

$$z_t = \begin{cases} \text{sign}(\varepsilon_{1t}) & \text{if NR is available on } t \\ 0 & \text{otherwise} \end{cases}$$

  and perform estimation and inference (e.g., Montiel-Olea, Stock & Watson 2021).

Questions

- An alternative approach to exploit narratives is the proxy (external) instrumental variable (IV) approach (Stock 2008, Mertens & Ravn 2013, Stock & Watson 2018).
- E.g., transform shock-sign restrictions into a proxy IV; for $t = 1, \ldots, T$,

$$z_t = \begin{cases} \text{sign}(\varepsilon_{1t}) & \text{if NR is available on } t \\ 0 & \text{otherwise} \end{cases}$$

  and perform estimation and inference (e.g., Montiel-Olea, Stock & Watson 2021).
- **Can the narrative proxy identify causal impulse response?**
- **Does it offer (frequentist) valid inference?**
- **How does it compare with the robust Bayes approach?**

Outline

1. Set up NR-SVAR models
2. Review: Bayesian/robust-Bayesian approaches to NR
3. Review: weak-proxy robust Anderson-Rubin (AR) confidence intervals (Montiel-Olea, Stock & Watson 2021)
4. New: Show that if the number of NRs is small, weak-proxy robust Anderson-Rubin confidence intervals do **not** have valid coverage even asymptotically ($T \to \infty$)
5. Monte Carlo study

Simple setting

Bivariate SVAR(0) + shock-sign NRs

- For $t = 1, \ldots, T$

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{A_0} \underbrace{\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}}_{y_t} = \underbrace{\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}}_{\varepsilon_t}, \quad \varepsilon_t \overset{iid}{\sim} N(0, I)$$

- Assume $1 \leq K \leq T$ number of shock-sign restrictions for ε_{1t} imposed for the first K periods, i.e., $\text{sign}(\epsilon_{1t})$ is observed for $t = 1, \ldots, K$.

- Shock-sign narrative proxy:

$$z_t = \begin{cases} \text{sign}(\varepsilon_{1t}) & \text{for } t = 1, \ldots, K \\ 0 & \text{for } t = K + 1, \ldots, T. \end{cases} \tag{1}$$

Bivariate example (SVAR(0))

$$Ay_t = \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N(0, I)$$

- Reduced form is $y_t \overset{iid}{\sim} N(0, \Sigma)$ with $\Sigma = A^{-1}(A^{-1})'$
- $\phi = \text{vech}(\Sigma_{tr})$ is the reduced-form parameter, where $\Sigma_{tr}\Sigma_{tr}' = \Sigma$ and

$$\Sigma_{tr} = \begin{bmatrix} \sigma_{11} & 0 \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Identification problem

- Multiple structural parameters compatible with reduced-form parameter:

  $$A^{-1} = \Sigma_{tr} Q$$

  where Q is an orthonormal matrix

  $$Q \in \left\{ \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \right\} \cup \left\{ \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} \right\}$$

  with $\theta \in [-\pi, \pi]$
- Object of interest η_{21} is the impulse-response of y_{2t} with respect to the first shock scaled by its own impulse response.

  $$\eta_{21} = \frac{(A^{-1})_{(2,1)}}{(A^{-1})_{(1,1)}} = \frac{\sigma_{21}}{\sigma_{11}} + \frac{\sigma_{22}}{\sigma_{11}} \tan\theta$$

Shock-sign NR

- Shock-sign restriction $\varepsilon_{1k} \geq 0 \Leftrightarrow \begin{pmatrix} 1 & 0 \end{pmatrix} Q'(\Sigma'_{tr})^{-1} y_k \geq 0$

$$\sigma_{22} y_{1k} \cos\theta + (\sigma_{11} y_{2k} - \sigma_{21} y_{1k}) \sin\theta \geq 0$$

- If $\sigma_{21} y_{1k} - \sigma_{11} y_{2k} < 0$ and $y_{1k} > 0$,

$$\theta \in \left[\arctan\left(\frac{\sigma_{22} y_{1k}}{\sigma_{21} y_{1k} - \sigma_{11} y_{2k}} \right), \arctan\left(\frac{\sigma_{22}}{\sigma_{21}} \right) \right].$$

- Mapping depends on y_k

Bayesian/robust Bayesian approach

Bayesian approach

- Write the shock-sign restrictions as $\{N(\theta, \phi, y_t) \geq 0\}_{t=1}^{K}$
- The likelihood based on sample $y^T = (y_1', \ldots, y_T')'$ and $\{N(\theta, \phi, y_t) \geq 0\}_{t=1}^{K}$ using unconditional likelihood:

$$p(y^T, \{N(\theta, \phi, y_t) \geq 0\}_{t=1}^{K} | \theta, \phi)$$

- Combine the likelihood with a prior for θ over $[-\pi, \pi]$
- Obtain the posterior for θ and η_{21}

Unconditional likelihood

- unconditional likelihood:

$$p(y^T, \{N(\theta,\phi,y_t) \geq 0\}_{t=1}^K | \theta,\phi) = \underbrace{\prod_{t=1}^T (2\pi)^{-1} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} y_t' \Sigma^{-1} y_t\right)}_{f(y^T|\phi)} \cdot \prod_{t=1}^K 1\{N(\theta,\phi,y_t) \geq 0\}$$

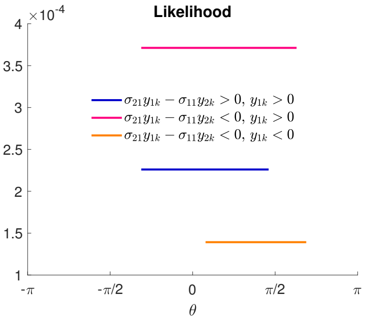
- Dependence of the likelihood on θ is only via $1\{N(\theta,\phi,y_t) \geq 0\}$, $t = 1,\ldots,K$
- NR truncates likelihood so fixing ϕ it is flat for θ satisfying NR and is zero otherwise, and points of truncation depend on $\{y_t\}_{t=1}^K$.

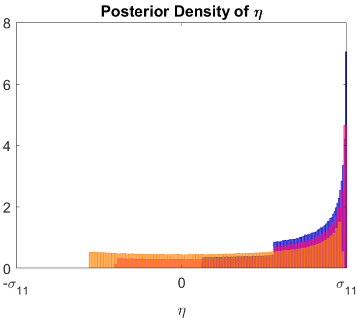
Nonstandard Features

SVARs under NR:

- GKR21 shows that the structural parameters are point-identified despite that a maximum likelihood estimator is not unique.
- Posterior inference becomes sensitive to a choice of prior for θ, similarly to what happens to Bayesian inference under traditional set-identifying restrictions (e.g., Baumeister & Hamilton, 2015)
- Posterior sensitivity analysis is important especially when “agnostic” or “non-informative” prior for θ is used.

Posterior for impulse response





Robust-Bayes Inference

- GKR21 proposes robust-Bayes inference for a parameter of interest $\eta = \eta(\theta, \phi)$.

Robust Credible region (GKR21):

1. Estimate reduced-form VAR to obtain the posterior for ϕ.
2. Draw ϕ from the posterior, and get the conditional identified set for η:

$$\begin{aligned} CIS_\eta(\phi | \{y_t\}_{t=1}^{K}) &\equiv \{\eta(\theta, \phi) : N(\theta, \phi, y_t) \geq 0, t = 1, \ldots, K\} \\ &= [\ell(\phi), u(\phi)] \end{aligned}$$

3. Robust credible region C_α with credibility α:

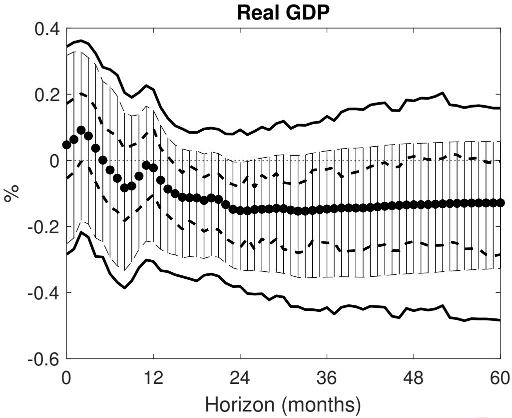
$$\widehat{Pr}([\ell(\phi), u(\phi)] \subset C_\alpha | y^T) = \alpha.$$

Frequentist properties of C_α

- Under regularity and fixed K, C_α attains correct frequentist coverage asymptotically as $T \to \infty$.
- For η_{21}, $CIS_{\eta_{21}}(\phi|\{y_t\}_{t=1}^K)$ can be unbounded depending on σ_{21} and $\{y_t\}_{t=1}^K$.
- C_α can be $\mathbb{R}$ if $\Pr(CIS_{\eta_{21}}(\phi|\{y_t\}_{t=1}^K) = \mathbb{R}|y^T) \geq 1-\alpha$.

- SVAR in AR18: 6 variables, 12 lags, monthly 1965-2007, η: output response to the unit s.d. monetary policy shock.
- Unconditional likelihood, NR: the Volcker shock being positive and larger than the other shocks in Oct. 1979

Figure: Impulse Responses to a Monetary Policy Shock



Proxy IV approach

Proxy IV approach

- **Main idea**: η_{21} can be identified by the linear IV Wald-estimand

  $$y_{2t} = \eta_{21} y_{1t} + \varepsilon_{2t},$$

  instrumenting y_{1t} by z_t satisfying $E(z_t \varepsilon_{1t}) \neq 0$ and $E(z_t \varepsilon_{2t}) = 0$.

- Shock-sign proxy IV

  $$z_t = \begin{cases} \text{sign}(\varepsilon_{1t}) & \text{for } t = 1, \ldots, K \\ 0 & \text{otherwise} \end{cases}$$

  satisfies both requirements for IV for $t = 1, \ldots, K$.

Narrative Proxy Identification

Lemma: NP identification

Let z_t be the shock-sign NP and define

$$\gamma_{1t} = T^{-1}\sum_{t=1}^{T} E[z_t y_{1t}] = T^{-1}\sum_{t=1}^{K} E[\text{sign}(\varepsilon_{1t})y_{1t}]$$

$$\gamma_{2t} = T^{-1}\sum_{t=1}^{T} E[z_t y_{2t}] = T^{-1}\sum_{t=1}^{K} E[\text{sign}(\varepsilon_{1t})y_{2t}].$$

Then, the impulse response of interest η_{21} can be identified by the Wald estimand:

$$\eta_{21} = \frac{\gamma_{2t}}{\gamma_{1t}} = \frac{\sum_{t=1}^{K} E[\text{sign}(\varepsilon_{1t})y_{2t}]}{\sum_{t=1}^{K} E[\text{sign}(\varepsilon_{1t})y_{1t}]}.$$

NP is a Weak IV

- If $K \leq \sqrt{T}$, $\gamma_{1t} \leq O(T^{-1/2})$ and NP is a weak instrument
- 2SLS estimator $\hat{\eta}_{21} = \frac{\sum_{t=1}^{K} \text{sign}(\varepsilon_{1t}) y_{2t}}{\sum_{t=1}^{K} \text{sign}(\varepsilon_{1t}) y_{1t}}$ suffers from the weak instrument problem
- $\hat{\eta}_{21}$ is biased, not consistent, the sampling distribution is far from normal, and can be outside of $CIS_{\eta}(\hat{\phi} | \{y_t\}_{t=1}^{K})$
- A standard solution to weak IV: weak-instrument robust inference, e.g., Anderson-Rubin (AR) confidence intervals

AR Confidence Intervals

- AR confidence intervals are constructed by inverting tests based on Wald statistics:

$$W_T(\eta_{21}) = \frac{T(\hat{\gamma}_2 - \eta_{21}\hat{\gamma}_1)^2}{\widehat{Var}(\sqrt{T}(\hat{\gamma}_2 - \eta_{21}\hat{\gamma}_1))},$$

where $\hat{\gamma}_1 = T^{-1}\sum_{t=1}^{T} z_t y_{1t}$ and $\hat{\gamma}_2 = T^{-1}\sum_{t=1}^{T} z_t y_{2t}$

AR Confidence Intervals

- AR confidence intervals are constructed by inverting tests based on Wald statistics:

$$W_T(\eta_{21}) = \frac{T(\hat{\gamma}_2 - \eta_{21}\hat{\gamma}_1)^2}{\widehat{Var}(\sqrt{T}(\hat{\gamma}_2 - \eta_{21}\hat{\gamma}_1))},$$

  where $\hat{\gamma}_1 = T^{-1}\sum_{t=1}^{T} z_t y_{1t}$ and $\hat{\gamma}_2 = T^{-1}\sum_{t=1}^{T} z_t y_{2t}$

- $W_T(\eta_{21}) \rightarrow_d \chi^2(1)$ under the null if

$$\sqrt{T}\begin{pmatrix}\hat{\gamma}_1 - \gamma_1 & \hat{\gamma}_2 - \gamma_2\end{pmatrix}' \rightarrow_d \mathcal{N}(0, \Omega)$$

  and $\widehat{Var}(\sqrt{T}(\hat{\gamma}_2 - \eta_{21}\hat{\gamma}_1))$ consistently estimates $\begin{pmatrix}-\eta_{21} & 1\end{pmatrix}\Omega\begin{pmatrix}-\eta_{21} & 1\end{pmatrix}'$.

Validity of AR is not guaranteed

- However, with K fixed, the conditions for asymptotic validity of AR confidence intervals fail

Validity of AR is not guaranteed

- However, with K fixed, the conditions for asymptotic validity of AR confidence intervals fail
- For any T, the numerator term of $W_T(\eta_{21})$ multiplied by $\sqrt{T}$

$$T^2(\hat{\gamma}_2 - \eta_{21}\hat{\gamma}_1)^2 = \left[\begin{pmatrix} -\eta_{21} & 1 \end{pmatrix} \hat{A}^{-1} \left(\sum_{t=1}^{K} \text{sign}(\varepsilon_{1t})\varepsilon_t \right)\right]^2$$

  and the denominator term of $W_T(\eta_{21})$ multiplied by $\sqrt{T}$ remains random even when $T \to \infty$.
- Hence, $W_T(\eta_{21})$ does not converge to $\chi^2(1)$, resulting in biased coverage of the AR confidence intervals.

Theorem

(i) With fixed K. $W_T(\eta_{21})$ does not converge to $\chi^2(1)$ asymptotically as $T \to \infty$.

(ii) If $K \to \infty$ as $T \to \infty$, $W_T(\eta_{21}) \to_d \chi^2(1)$ as $T \to \infty$.

Size (coverage) distortion

CDFs of $W_T(\eta_{21})$ vs. $\chi^2(1)$

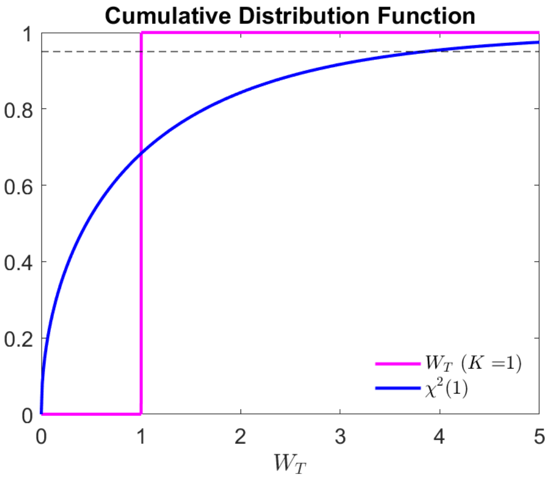


Figure: $K = 1$, $\eta_{21} = 0.4$, $\text{vech}(\Sigma_{tr}) \approx (0.7, -1.1, 1.4)$

Size (coverage) distortion

CDFs of $W_T(\eta_{21})$ vs. $\chi^2(1)$

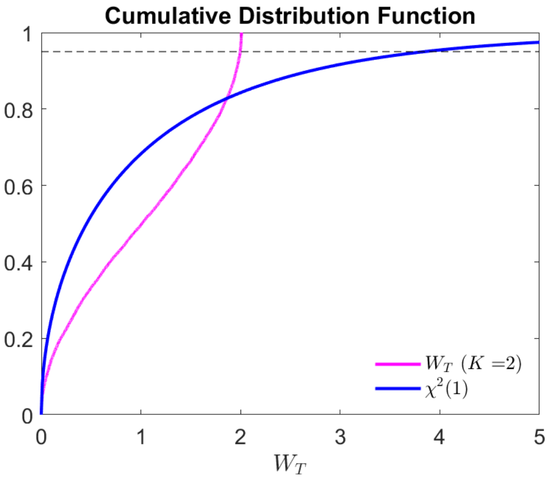


Figure: $K = 2$, $\eta_{21} = 0.4$, $\text{vech}(\Sigma_{tr}) \approx (0.7, -1.1, 1.4)$

Size (coverage) distortion

CDFs of $W_T(\eta_{21})$ vs. $\chi^2(1)$

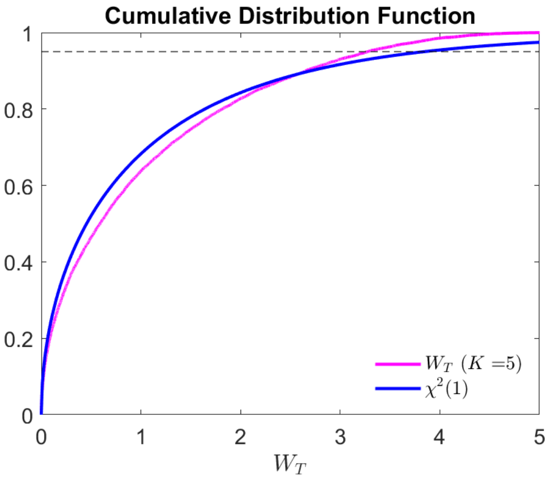


Figure: $K = 5$, $\eta_{21} = 0.4$, $\text{vech}(\Sigma_{tr}) \approx (0.7, -1.1, 1.4)$

Size (coverage) distortion

CDFs of $W_T(\eta_{21})$ vs. $\chi^2(1)$

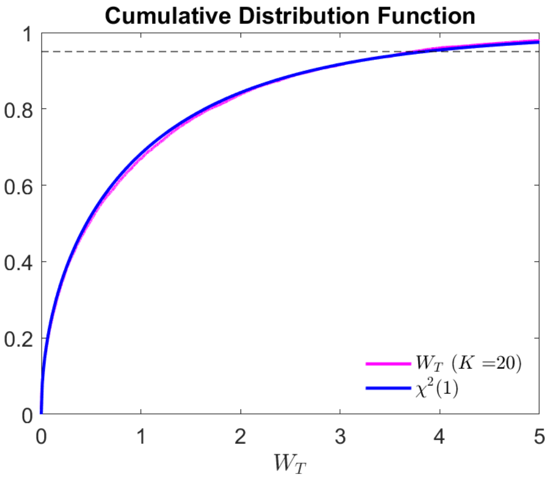


Figure: $K = 20$, $\eta_{21} = 0.4$, $\text{vech}(\Sigma_{tr}) \approx (0.7, -1.1, 1.4)$

Summary

- With fixed K, the null distribution of $W_T(\eta_{21})$ is no longer $\chi^2(1)$ even asymptotically ($T \to \infty$).
- AR confidence intervals can undercover or overcover depending on the true data generating process and confidence level.

Summary

- With fixed K, the null distribution of $W_T(\eta_{21})$ is no longer $\chi^2(1)$ even asymptotically ($T \to \infty$).
- AR confidence intervals can undercover or overcover depending on the true data generating process and confidence level.
- Can we modify? Given η_{21} and $\hat{\Sigma}_{tr}$, we could simulate

$$T^2(\hat{\gamma}_2 - \eta_{21}\hat{\gamma}_1)^2 = \left[\begin{pmatrix} -\eta_{21} & 1 \end{pmatrix} \hat{A}^{-1} \left(\sum_{t=1}^{K} \text{sign}(\varepsilon_{1t})\varepsilon_t \right)\right]^2$$

- As K becomes moderate to large, the null distribution of $W_T(\eta_{21})$ can be well approximated by $\chi^2(1)$.

Monte Carlo (in progress)

- $T = 500$, $K = 1, 2, 5, 10, 20, 100$.
- $A^{-1} = \begin{pmatrix} 0.5 & -0.5 \\ 0.2 & 1.8 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 0.5 & -0.8 \\ -0.8 & 3.28 \end{pmatrix}$.
- Confidence level, 95% and 68%
- We report Coverage, Pr(unbounded CI), and Median length.

$\alpha = 0.95$	coverage		Pr. unbounded		Med. length	
K	AR	GKR21	AR	GKR21	AR	GKR21
1	1	1	1	0.78	∞	∞
5	0.98	1	0.89	0.29	∞	9.86
10	0.96	1	0.56	0.08	∞	6.95
20	0.95	1	0.17	0.01	8.26	6.11
100	0.95	0.99	0	0	2.13	5.54

Table: Monte Carlo: $T = 500$ and $\alpha = 0.95$

$\alpha = 0.68$ K	coverage AR	Pr. unbounded AR	Med. length AR
1	0	1	0
2	0.50	0.37	7.58
5	0.63	0.27	6.14
10	0.66	0.11	3.68
20	0.68	0.02	2.39

Table: Monte Carlo: $T = 500$ and $\alpha = 0.68$

Conclusion

- NR-SVAR is becoming popular, while econometric methods for it are less studied.
- We showed that the standard AR confidence intervals applied to a shock-sign proxy IV are not guaranteed to be valid (cf. robust Bayes credible region).
- According to Monte Carlo, the coverage of AR confidence intervals becomes as desired when $K \geq 20$, and they tend to be shorter than the robust Bayes credible region.

Conclusion

- NR-SVAR is becoming popular, while econometric methods for it are less studied.
- We showed that the standard AR confidence intervals applied to a shock-sign proxy IV are not guaranteed to be valid (cf. robust Bayes credible region).
- According to Monte Carlo, the coverage of AR confidence intervals becomes as desired when $K \geq 20$, and they tend to be shorter than the robust Bayes credible region.
- Left for future: other type of NRs? Other approach to frequentist inference? How to model the mechanism of generating NRs? Applications to microeconometrics?