

Discussion of “Narrative Restrictions and Proxies” by Raffaella Giacomini, Toru Kitagawa, and Matthew Read*

Mikkel Plagborg-Møller
Princeton University

February 25, 2022

I am grateful for the chance to discuss this characteristically insightful paper by Giacomini, Kitagawa, and Read (henceforth GKR). Since the seminal contribution of [Antolín-Díaz & Rubio-Ramírez \(2018\)](#), narrative restrictions have rapidly become one of the go-to tools for sharpening causal inference in SVAR analysis. [Giacomini, Kitagawa & Read \(2021\)](#) contributed greatly to our understanding of the role of subjective prior beliefs and the appropriate form of the likelihood function when exploiting such narrative information. In the new paper that is the topic of this discussion, GKR compare their preferred prior-robust Bayesian inference procedure with an alternative approach that constructs categorical proxy variables from the narrative information and uses these to estimate impulse responses via instrumental variable (IV) regressions. GKR argue that the proxy approach will likely suffer from weak IV problems when we only have narrative restrictions for a few time periods, as is often the case in practice. To add insult to injury, this cannot be addressed using existing techniques for weak-IV-robust inference in SVARs ([Montiel Olea, Stock & Watson, 2021](#)).

In the following I will make two points. First, the proxy approach to exploiting narrative information has several appealing robustness properties relative to the likelihood approaches of [Antolín-Díaz & Rubio-Ramírez \(2018\)](#) and [Giacomini et al. \(2021\)](#): The proxy approach allows the narrative signals to be imperfect and arrive non-randomly, and furthermore, the economic shocks are allowed to be non-invertible (also known as non-fundamental). Second, the weak IV problem that GKR discuss can be overcome by using procedures designed for small samples, such as permutation tests.

*Email: mikkelpm@princeton.edu. Presented at the JBES Session of the 2022 North American Winter Meeting of the Econometric Society. I am grateful for comments from Lutz Kilian, Toru Kitagawa, Pepe Montiel Olea, Eric Qian, and Juan Rubio-Ramírez. I acknowledge that this material is based upon work supported by the NSF under Grant #1851665. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF.

1 Robustness of the proxy approach

To explain why I believe the proxy approach to narrative identification to be particularly robust, I will present the narrative SVAR model a bit differently from GKR. The model remains mathematically the same, however. Consider a static SVAR model for the two observed variables (y_{1t}, y_{2t}) :

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \quad t = 1, \dots, T. \quad (1)$$

The two unobserved shocks $(\varepsilon_{1t}, \varepsilon_{2t})$ are assumed to be i.i.d. mean zero, variance 1, and mutually independent of each other. I normalize $\Theta_{11}, \Theta_{22} > 0$. Like GKR, I focus on doing inference on the relative impulse response parameter $\tilde{\eta}_{21} \equiv \Theta_{21}/\Theta_{11}$, which is the response of y_{2t} with respect to an impulse in ε_{1t} that raises y_{1t} by 1 unit. Without further information, the identified set for $\tilde{\eta}_{21}$ is large (in fact, the entire real line).

To sharpen identification, the econometrician seeks to exploit (what I will call) *narrative signals*. For simplicity, I restrict attention to signals about the shock of interest, ε_{1t} . The narrative signals are encoded in a variable $Z_t \in \{-1, 0, 1\}$, $t = 1, \dots, T$. We interpret $Z_t = -1$ to mean that the shock ε_{1t} is believed to be negative at time t , while $Z_t = 1$ signifies a belief that the shock was positive. The value $Z_t = 0$ means that we have no narrative information about ε_{1t} in that period. Note that information about shock signs is merely one of the types of narrative information considered by [Antolín-Díaz & Rubio-Ramírez \(2018\)](#).

1.1 Inference when signals are perfect

The Bayesian inference procedures of [Antolín-Díaz & Rubio-Ramírez \(2018\)](#) and [Giacomini et al. \(2021\)](#) rely on the assumption that the narrative signals are *perfect*:

$$Z_t = \text{sign}(\varepsilon_{1t}) \quad \text{whenever } Z_t \neq 0. \quad (2)$$

Here I define $\text{sign}(x) = 1$ when $x > 0$, $\text{sign}(x) = -1$ when $x < 0$, and $\text{sign}(0) = 0$.

Under this assumption, it follows from the SVAR model (1) that

$$\text{sign} \left(\frac{\Theta_{22}y_{1t} - \Theta_{12}y_{2t}}{\Theta_{11}\Theta_{22} - \Theta_{12}\Theta_{21}} \right) = Z_t \quad \text{whenever } Z_t \neq 0.$$

This amounts to a set of data-dependent inequality restrictions on the impulse response

parameters. Evidently, these inequalities can potentially be very helpful in restricting the values of $\tilde{\eta}_{21} = \Theta_{21}/\Theta_{11}$ that are consistent with the data. The above-mentioned Bayesian inference procedures impose these inequality restrictions dogmatically.

In my view, the explicit and implicit assumptions imposed by these inference procedures are quite strong, for three reasons.

1. The assumption (2) that signals are perfect requires there to be 100% certainty about the exact dating of the known positive or negative shocks. If we got 1,000 economists together in a room, would they all agree that the contractionary “Volcker shock” was revealed to the world in October 1979, rather than September, say?
2. The functional form of the likelihood functions used by Antolín-Díaz & Rubio-Ramírez (2018) and Giacomini et al. (2021) appear to assume that the narrative signals arrive randomly over time (i.e., the event $\{Z_t \neq 0\}$ is independent of $(\varepsilon_{1t}, \varepsilon_{2t})$), as noted in Section 3.4.2 of the latter paper.¹ Thus, the econometrician’s knowledge about the sign of the “Volcker shock” is ascribed to serendipity rather than the fact that this shock was particularly large in magnitude.
3. All SVAR models, of the form (1) but with more variables/shocks and added lag dynamics, assume that the structural shocks $(\varepsilon_{1t}, \varepsilon_{2t})$ are *invertible*, i.e., spanned by current and past values of the data (y_{1t}, y_{2t}) . This assumption is questionable in certain applications, see for example Chahrour & Jurado (2022).

I will now argue that these assumptions can all be relaxed by switching to the proxy approach to narrative identification.

1.2 Robust inference

The proxy approach allows signals to be imperfect and arrive non-randomly. Specifically, consider replacing the condition (2) with the weaker assumption that the signals arise from the non-parametric signal generation model

$$Z_t = F(\varepsilon_{1t}, u_t), \tag{3}$$

¹For example, suppose that the signal $Z_t = 1$ is in fact received if and only if the shock ε_{1t} is not only positive, but also large, e.g., $\varepsilon_{1t} \geq 2$. Then the correct likelihood function would truncate at 2 instead of 0.

where $F: \mathbb{R}^2 \rightarrow \{-1, 0, 1\}$ is an unknown function, and u_t is a random variable that is independent of $(\varepsilon_{1t}, \varepsilon_{2t})$ at all leads and lags. I interpret u_t as non-classical measurement error. This model allows for a wide range of (unknown) signal generating mechanisms.² The key restriction imposed by the model (3) is that the narrative signals are not contaminated by the nuisance shock ε_{2t} . This restriction is also imposed by the previously-mentioned Bayesian inference procedures, as discussed above.

For the imperfect signals to contain some useful information, I require that they are not too inaccurate on average, in the sense

$$\text{Cov}(Z_t, \varepsilon_{1t}) > 0. \quad (4)$$

A sufficient (but not necessary) condition is that $Z_t = \text{sign}(E(\varepsilon_{1t} | Z_t))$ whenever $Z_t \neq 0$.

Under assumptions (3)–(4), the relative impulse response $\tilde{\eta}_{21}$ can be consistently estimated from an IV regression of y_{2t} on y_{1t} , using the proxy Z_t as an IV. This follows because simple manipulation of the SVAR model (1) yields

$$y_{2t} = \tilde{\eta}_{21} y_{1t} + \left(\Theta_{22} - \frac{\Theta_{12}\Theta_{21}}{\Theta_{11}} \right) \varepsilon_{2t}, \quad (5)$$

and my assumptions immediately imply that Z_t is an exogenous and relevant IV in this regression equation. The use of categorical “event proxies” like Z_t in macroeconometrics is a practice that in spirit dates back at least to [Romer & Romer \(1989\)](#), [Hoover & Perez \(1994\)](#), and [Hamilton \(2003\)](#).³

To summarize, the proxy approach to narrative identification is able to dispense with some of the strong assumptions made in the previous subsection. First of all, we can allow the signals Z_t to be imperfect (i.e., contaminated by measurement error) and to arrive non-randomly (i.e., in a way that depends on the magnitude of ε_{1t} in addition to its sign). Moreover, in a dynamic context we may estimate a Local Projection version of the above static IV regression, which allows the shocks $(\varepsilon_{1t}, \varepsilon_{2t})$ to be possibly non-invertible, as shown in a general setting by [Stock & Watson \(2018\)](#) (and mentioned by GKR in a footnote).⁴

²One example is $Z_t = \text{sign}(\varepsilon_{1t} + u_t) \mathbb{1}(|\varepsilon_{1t}| \geq 2)$, whereby signals are observed only when the shock is particularly large in magnitude, and even then, the sign of the shock may be randomly misclassified.

³See [Budnik & Rünstler \(2020\)](#) and [Boer & Lütkepohl \(2021\)](#) for recent econometric analyses.

⁴Equivalently, we could estimate impulse responses from a recursively identified SVAR with Z_t included and ordered first, cf. [Plagborg-Møller & Wolf \(2021, Section 3.3\)](#).

2 Inference with a weak proxy

GKR point out a key challenge in applying the proxy approach in practice: Because we typically only have narrative information pertaining to a small number K of time periods, the proxy Z_t is likely a weak IV. Unfortunately, the off-the-shelf weak-IV-robust SVAR inference procedures of [Montiel Olea et al. \(2021\)](#) fail under asymptotics where K is held fixed as $T \rightarrow \infty$. This is because sample moments involving Z_t are not asymptotically normal at the usual $\sqrt{T}$ convergence rate if Z_t equals zero in all but a finite number of periods. GKR propose a modified inference procedure that is valid when the shocks are Gaussian, but one would like to avoid relying on such strong distributional assumptions.

Does the weak IV issue mean that we must give up on the robustness afforded by the proxy approach? Not necessarily. The weak proxy issue is essentially a small-sample problem, as we are trying to learn economic structure from identifying information that pertains to a small number K of periods. We should therefore apply procedures that are geared specifically towards small samples.

One such small-sample weak-IV-robust inference procedure is the *permutation Anderson-Rubin test* of [Imbens & Rosenbaum \(2005\)](#). Consider testing the null hypothesis $H_0: \tilde{\eta}_{21} = \bar{\eta}$, where $\bar{\eta}$ is some particular value (as usual, we can invert this test to obtain a confidence interval). Under the null, equation (5) and assumption (3) imply that $(y_{2t} - \bar{\eta}y_{1t})$ is independent of Z_t . We may therefore test the hypothesis by applying a Fisher permutation test of independence. For example:

1. Compute $|\widehat{\text{Corr}}(y_{2t} - \bar{\eta}y_{1t}, Z_t)|$.⁵
2. Compute the same statistic over all possible permutations of the IV data points $Z_1, \dots, Z_T$.
3. Reject H_0 if the original statistic from Step 1 exceeds the 95th percentile of the permutation distribution from Step 2.

By standard arguments, this test has size *exactly* equal to 5% in finite samples, regardless of how large or small T and K are.⁶ Note that this test does not require the shocks to be

⁵Other test statistics are possible, such as the Kolmogorov-Smirnov statistic for comparing the empirical distribution of $(y_{2t} - \bar{\eta}y_{1t})$ in the subsample with $Z_t = 1$ against the subsample with $Z_t = -1$.

⁶In a dynamic version of the SVAR model (1) we would first need to estimate the lag coefficients to impute the VAR residuals. The estimation error would then cause the size of the permutation test to differ from 5% in finite samples. However, the size would converge to 5% as $T \rightarrow \infty$, for any finite value of K , under standard regularity conditions.

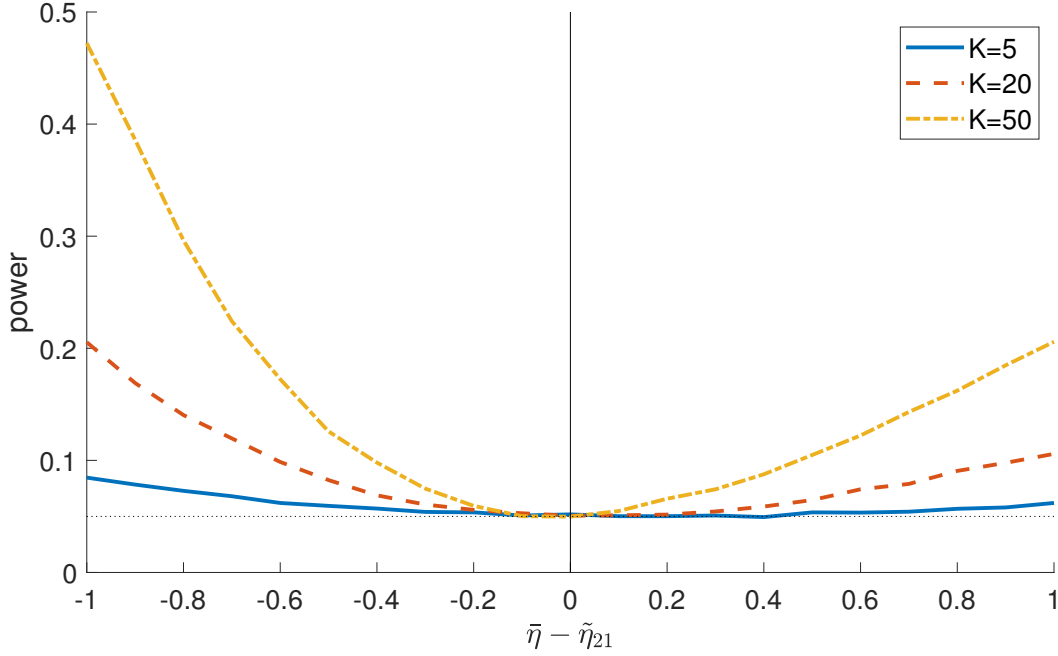


Figure 1: Rejection frequency of Anderson-Rubin permutation test as a function of the difference between the hypothesized and true values of $\tilde{\eta}_{21}$. Different curves correspond to different numbers $K \in \{5, 20, 50\}$ of narrative signals. Horizontal dotted line marks the nominal significance level of 5%. $T = 500$; 5,000 simulations per DGP; 1,000 random permutations per test.

Gaussian, only mutually independent.

Figure 1 illustrates the size and power properties of the permutation Anderson-Rubin test. The DGP is essentially the same as in GKR’s simulation study.⁷ Evidently, the size of the test equals 5% regardless of K , consistent with theory. The power increases with the number K of narrative signals received. In this particular DGP, the power is close to trivial when only 5 signals are observed, even for large violations of the null hypothesis (the true value of $\tilde{\eta}_{21}$ equals 0.4). The power is modest even with 50 signals, though sufficient to stand a decent chance at rejecting large negative values for the relative impulse response. I do not intend to claim that these findings about power will generalize to other DGPs, as one would expect the power to depend on the share of the variance of the data that is explained by the shock of interest.

⁷The only difference is that I do not generate the K first shocks from a truncated normal distribution. Instead, all shocks have an unconditional normal distribution, and I set $Z_t = 1$ for the first K periods where $\varepsilon_{1t} > 0$, while $Z_t = 0$ in all other periods.

3 Conclusion

GKR have done the profession a service by drawing attention to the challenges involved in doing robust structural inference when narrative information is available for a relatively small number of time periods. I have argued that this challenge can be addressed using weak-IV-robust procedures that are appropriate in small samples, such as permutation tests. In this way, we can continue to benefit from the robustness of the proxy approach in the face of signal imperfections and shock non-invertibility. The statistical power may be modest in realistic DGPs, but this could be viewed as a natural limitation of relying on identifying information that only pertains to a few time periods.

This area seems ripe for further research. First, one could investigate the gain in power from combining the proxy approach to narrative restrictions with other types of identifying information. Second, while I have focused on narrative information about shock signs, it would be interesting to extend the analysis to information about historical decompositions (Antolín-Díaz & Rubio-Ramírez, 2018, pp. 2807–2809). Third, more work is needed to characterize the power properties of the permutation Anderson-Rubin test, and to adapt this test to a dynamic Local Projection context. Finally, it would be useful to relax the assumption that the structural shocks are mutually independent, in order to allow for shared volatility factors (which are also ruled out by other approaches to narrative identification, to my knowledge).

References

- Antolín-Díaz, J. & Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for SVARs. *American Economic Review*, 108(10), 2802–2829.
- Boer, L. & Lütkepohl, H. (2021). Qualitative versus quantitative external information for proxy vector autoregressive analysis. DIW Berlin Discussion Paper 1940.
- Budnik, K. & Rünstler, G. (2020). Identifying SVARs from sparse narrative instruments. European Central Bank Working Paper 2353.
- Chahrour, R. & Jurado, K. (2022). Recoverability and expectations-driven fluctuations. *Review of Economic Studies*, 89(1), 214–239.
- Giacomini, R., Kitagawa, T., & Read, M. (2021). Identification and inference under narrative restrictions. arXiv:2102.06456.
- Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, 113(2), 363–398.
- Hoover, K. D. & Perez, S. J. (1994). Post hoc ergo propter once more: An evaluation of ‘Does monetary policy matter?’ in the spirit of James Tobin. *Journal of Monetary Economics*, 34(1), 47–74.
- Imbens, G. W. & Rosenbaum, P. R. (2005). Robust, accurate confidence intervals with a weak instrument: Quarter of birth and education. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 168(1), 109–126.
- Montiel Olea, J. L., Stock, J. H., & Watson, M. W. (2021). Inference in Structural Vector Autoregressions identified with an external instrument. *Journal of Econometrics*, 225(1), 74–87. Themed Issue: Vector Autoregressions.
- Plagborg-Møller, M. & Wolf, C. K. (2021). Local projections and VARs estimate the same impulse responses. *Econometrica*, 89(2), 955–980.
- Romer, C. D. & Romer, D. H. (1989). Does monetary policy matter? A new test in the spirit of Friedman and Schwartz. In O. Blanchard & S. Fischer (Eds.), *NBER Macroeconomics Annual*, volume 4 (pp. 121–170). University of Chicago Press.
- Stock, J. H. & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *Economic Journal*, 128(610), 917–948.