Local Projection Inference in High Dimensions

Local Projection Inference in High Dimensions

Robert Adamek[*][a], Stephan Smeekes[a], and Ines Wilms[a]

[a]Department of Quantitative Economics, Maastricht University, The Netherlands

September 8, 2022

Abstract

In this paper, we estimate impulse responses by local projections in high-dimensional settings. We use the desparsified (de-biased) lasso to estimate the high-dimensional local projections, while leaving the impulse response parameter of interest unpenalized. We establish the uniform asymptotic normality of the proposed estimator under general conditions. Finally, we demonstrate small sample performance through a simulation study and consider two canonical applications in macroeconomic research on monetary policy and government spending.

Keywords: high-dimensional data, honest inference, impulse responses, lasso, time series

1 Introduction

In this paper, we develop a simple approach for conducting valid inference on impulse responses in high-dimensional settings. We do so by pairing the estimation framework of local projections (LPs) with the inference framework of the desparsified (or debiased) lasso. Since their introduction by Jordà (2005), LPs have been widely used in macroeconomic research for studying the dynamic propagation of shocks through impulse response analysis, see e.g. Angrist et al. (2018), Ramey and Zubairy (2018), Ramey (2016) and Stock and Watson (2018). As such, they have become an increasingly used alternative to structural Vector AutoRegressions (SVAR) pioneered by Sims (1980), see for instance Ramey (2016) or Kilian and Lütkepohl (2017) for an overview on this work.

The distinct feature of LPs compared to SVARs is that only univariate regressions have to be estimated, as compared to estimating the whole system needed for the SVAR. This is particularly useful where system estimation is complicated, such as in the presence of nonlinearities or when the dimension of the system grows. Plagborg-Møller and Wolf (2021) established the equivalence between SVARs and LPs in population, showing that the underlying impulse response estimands

---

[*]Present Address: Department of Economics and Business Economics, Aarhus University, Fuglesangs All´e 4, 8210 Aarhus V, Denmark.

are the same for both, and emphasizing that identification strategies typically used for SVARs can equivalently be implemented with LPs, and vice versa. As such, the choice between SVAR or LP estimation of impulse responses becomes a finite-sample consideration, where the dimensionality of the problem may play a large role in this choice.

We consider high-dimensional local projections (HDLPs) in a general time series framework where the number of regressors can grow faster than the sample size. Impulse response analysis quickly becomes high-dimensional in macroeconomic research. Even when considering impulse response analysis with few variables, the number of regressors in LPs or SVARs is often large due to the common practices of including many lags to control for autocorrelation (see e.g., Bernanke and Mihov, 1998, Romer and Romer, 2004 and Sims and Zha, 2006) or to robustify against (near) unit roots via lag augmentation as in Montiel Olea and Plagborg-Møller (2021). Similarly, additional regressors are often introduced to capture seasonal patterns in impulse responses (see e.g., quarter-dependent coefficients used in Blanchard and Perotti, 2002), or to permit nonlinearities to produce state-dependent impulse responses (Koop et al., 1996; Ramey and Zubairy, 2018).

Additionally, one might be interested in impulse response analysis with many variables. Motivations of the latter include, amongst others, avoidance of contaminated measurements of policy innovations or informational deficiency with impulse response estimates that are distorted by omitted variable bias, the ability to use several observable measures of difficult-to-measure variables in theoretical models, or opportunities for researchers and policy makers to investigate the impact of shocks on a larger, possibly more disaggregated set of variables they care about (see e.g., Bernanke et al., 2005, Forni and Gambetti, 2014, Stock and Watson, 2016 and Kilian and Lütkepohl, 2017, Ch. 16).

Regular estimation approaches, however, become infeasible for high-dimensional LPs or SVARs. Traditional approaches to handle the high-dimensionality include modelling commonalities between variables, such as through the factor-augmented VAR (Bernanke et al., 2005), dynamic factor model (Forni et al., 2009; Stock and Watson, 2016) or – in the case of a panel structure – global VAR (Chudik and Pesaran, 2016); as well as through Bayesian shrinkage methods (Bańbura et al., 2010; Chan, 2020). Recently, sparse shrinkage methods such as the lasso have gained considerable popularity in econometrics for policy evaluation (Belloni et al., 2014) as well as (macroeconomic) time series analysis (Kock et al., 2020). Their use in structural impulse response analysis however has only very scarcely been explored so far.

While a multitude of methods and theoretical results now exist for the estimation of sparse VAR models – see e.g., Basu and Michailidis (2015); Kock and Callot (2015); Masini et al. (2022) and

the references cited therein – inference on impulse responses is complicated by two issues. First, sparse estimation techniques such as the lasso perform model selection, which induces issues with non-uniformity of limit results if this selection is ignored (Leeb and Pötscher, 2005). Second, while several methods such as orthogonalization (Belloni et al., 2014) and debiasing, or desparsifying the lasso (van de Geer et al., 2014; Javanmard and Montanari, 2014) have been proposed to yield uniformly valid (or 'honest') post-selection inference, the impulse response parameters are complex nonlinear functions of *all* VAR parameters. This severely limits the applicability of existing post-selection inference methods which are typically designed for (relatively) low-dimensional parameters of interest that can directly be estimated. Indeed, to our knowledge, impulse response analysis in sparse HD-SVARs is only considered in Krampe et al. (2022), who construct a complex multi-step algorithm to overcome these complications. Instead, by casting the problem in the LP framework, we reduce the impulse response parameter(s) to a (directly estimable) low-dimensional object in the presence of high-dimensional nuisance parameters, which makes the standard post-selection tools available.

In this paper, we develop HDLP inference for impulse responses based on the desparsified lasso of van de Geer et al. (2014) and its time series extension in Adamek et al. (2022b). We tailor the approach of Adamek et al. (2022b) specifically to high-dimensional LPs consisting of a small number of parameters of interest – typically the dynamic response of certain variable to a shock at a given horizon – and many controls. In particular, we modify their approach by leaving the parameter of interest unpenalized during the estimation procedure to ensure it does not suffer from penalization bias. Such a setting is of more general relevance for treatment effect models consisting of a small number of variables whose effects are of interest combined with a large set of controls. We theoretically show that the combination of few unpenalized parameters with many penalized ones does not affect the asymptotic behaviour of the desparsified lasso; it remains asymptotically normal. Importantly, we find considerable gains in terms of better coverage rates over the standard desparsified lasso in finite samples.

We consider two canonical macroeconomic applications and demonstrate the performance of the proposed desparsified-lasso based for HDLPs in recovering structural impulse responses. Specifically, we first extend the work of Bernanke et al. (2005) on macroeconomic responses to a shock in monetary policy to a HDLP setting which additionally allows for differing responses during the zero lower bound state. Second, we consider the work by Ramey and Zubairy (2018) on state-dependent impulse responses to a shock in government spending and extend it to a HDLP specification with more lags as well as a more complex state-dependent HDLP with interaction between different

state variables. In addition, an implementation of our work is available in the R package `desla` (Adamek et al., 2022a), which offers both the methods of Adamek et al. (2022b) as well as the high-dimensional local projections proposed here.

Section 2 introduces the HDLP specification with the corresponding desparsified lasso-based estimation and inference procedure. We present our main result on the asymptotic normality of the desparsified lasso. Section 3 contains a simulation study that examines the small sample performance of the proposed desparsified lasso. Section 4 presents the results on the two macroeconomic applications. Section 5 concludes.

A word on notation. For any N dimensional vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|_r = \left(\sum\limits_{i=1}^{N} |x_i|^r \right)^{1/r}$ denotes the l_r-norm, with the familiar convention that $\|\boldsymbol{x}\|_0 = \sum_i 1(|x_i| > 0)$. Depending on the context, $\sim$ denotes equivalence in order of magnitude of sequences, or equivalence in distribution.

2 High-dimensional Local Projections

Consider the local projection regression

$$y_{t+h} = \phi_h x_t + \boldsymbol{\eta}_h' \boldsymbol{x}_{s,t} + \sum_{k=1}^{K} \boldsymbol{\delta}_{h,k}' \boldsymbol{z}_{t-k} + u_{h,t}, \quad h = 0, 1, \dots, h_{\max}, \tag{1}$$

where $\phi_h, \boldsymbol{\eta}_h, \boldsymbol{\delta}_{h,k}$ are the projection parameters, the set of variables $\boldsymbol{z}_t = \left(\boldsymbol{x}_{s,t}', x_t, y_t, \boldsymbol{x}_{f,t}' \right)'$ consist of the shock variable x_t, the response y_t and the vectors of control variables $\boldsymbol{x}_{s,t} \in \mathbb{R}^{n_s}$, $\boldsymbol{x}_{f,t} \in \mathbb{R}^{n_f}$, and $u_{h,t}$ is the projection error.[1] Our single parameter of interest is ϕ_h, the dynamic response at horizon h of y_t after an impulse in x_t. The LP impulse response function of y_t with respect to x_t is given by $\{\phi_h\}_{h \geq 0}$ and can be obtained by estimating eq. (1) for $h = 0, 1, \dots, h_{\max}$. We propose a desparsified lasso estimator to obtain point-wise confidence intervals for ϕ_h at each horizon h.

For structural interpretation purposes, the partition of variables in $\boldsymbol{z}_t$ requires care. For simplicity and given the setup of our applications, we focus here on (partially) recursively identified impulse responses. Plagborg-Møller and Wolf (2021) provide detailed discussions on how other identification schemes, such as sign restrictions, can be imposed while estimating eq. (1). In a partially recursively identified system, the variable x_t can be thought of as being predetermined with respect to y_t, the "slow" variables in $\boldsymbol{x}_{s,t}$ as being predetermined with respect to $\{x_t, y_t\}$ and finally the "fast" variables in $\boldsymbol{x}_{f,t}$ can react to the shock variable x_t within the same observation period and should only enter the equation with a lag. To permit causal interpretation of the impulse responses, structural assumptions about the Data Generating Process (DGP) are needed. Example

---

[1]We omit the constant in our definition, since we generally demean the data when using the desparsified lasso.

1 considers identification via a Structural Vector Moving Average model (taken from Assumption 3 in Plagborg-Møller and Wolf, 2021) as an example.

Example 1 (SVMA)**.** Consider the Structural Vector Moving Average (SVMA)

$$\underset{n_z \times 1}{\boldsymbol{z}_t} = \boldsymbol{\mu} + \underset{n_z \times n_\epsilon}{\boldsymbol{A}(L)}\ \underset{n_\epsilon \times 1}{\boldsymbol{\epsilon}_t}\ , \quad \boldsymbol{A}(L) = \sum_{k=0}^{\infty} \underset{n_z \times n_\epsilon}{\boldsymbol{A}_k}\ L^k, \quad \boldsymbol{\epsilon}_t \overset{i.i.d.}{\sim} N(\boldsymbol{0}, \boldsymbol{I}), \tag{2}$$

where $\boldsymbol{\epsilon}_t$ is a vector of structural shocks, L is the lag operator, $\{\boldsymbol{A}_k\}_k$ is absolutely summable, and $\boldsymbol{A}(x)$ has full row rank for all complex scalars x on the unit circle. Under this assumption, the errors $u_{h,t}$ from eq. (1) are generally autocorrelated due to the leads of y_t on the left-hand-side. It is therefore important to estimate the LPs with a procedure that is robust to autocorrelation, as will be discussed in Section 2.1.

While the LP in eq. (1) has a single parameter of interest, it can easily be extended to allow for LPs with a small number of parameters of interest, see Example 2.

Example 2 (State-Dependent LP)**.** Consider the nonlinear state-dependent model of Ramey and Zubairy (2018):

$$
\begin{aligned}
y_{t+h} = I_{t-1}&\left[\alpha_h + \phi_{A,h}x_t + \sum_{k=1}^{K} \boldsymbol{\delta}'_{A,h,k}\boldsymbol{z}_{t-k}\right] \\
&+ (1 - I_{t-1})\left[\phi_{B,h}x_t + \sum_{k=1}^{K} \boldsymbol{\delta}'_{B,h,k}\boldsymbol{z}_{t-k}\right] + u_{h,t}.
\end{aligned}
\tag{3}
$$

Here, $\phi_{A,h}$ and $\phi_{B,h}$ are the two responses of y_t after an impulse in x_t, corresponding to two different states distinguished by the dummy variable I_t.

To generally accommodate LP settings with a small number of parameters of interest and a large number of controls, we will use the more general notation

$$y_t = \underset{1 \times S}{\boldsymbol{x}'_{\mathcal{S},t}}\underset{S \times 1}{\boldsymbol{\beta}_{\mathcal{S}}} + \underset{1 \times (N-S)}{\boldsymbol{x}'_{-\mathcal{S},t}}\underset{(N-S) \times 1}{\boldsymbol{\beta}_{-\mathcal{S}}} + u_t, \quad t = 1, \ldots, T, \tag{4}$$

and in matrix form

$$\underset{T \times 1}{\boldsymbol{y}} = \underset{T \times S}{\boldsymbol{X}_{\mathcal{S}}}\boldsymbol{\beta}_{\mathcal{S}} + \underset{T \times (N-S)}{\boldsymbol{X}_{-\mathcal{S}}}\boldsymbol{\beta}_{-\mathcal{S}} + \underset{T \times 1}{\boldsymbol{u}}.$$

We hereby separate the parameter vector $\underset{N \times 1}{\boldsymbol{\beta}} = [\boldsymbol{\beta}'_{\mathcal{S}}, \boldsymbol{\beta}'_{-\mathcal{S}}]'$ into two groups. The first group consists of the low-dimensional set of parameters of interest $\boldsymbol{\beta}_{\mathcal{S}}$, corresponding to the variables $\boldsymbol{x}_{\mathcal{S},t}$, which will be unpenalized during the estimation procedure to ensure they are not affected by penalization bias. The second group consists of the large set of parameters $\boldsymbol{\beta}_{-\mathcal{S}}$, corresponding to control variables $\boldsymbol{x}_{-\mathcal{S},t}$, which will be penalized during the estimation procedure. We hereby use $\mathcal{S}$

to denote the set that indexes the S variables of interest. Without loss of generality, we order these variables first in $\underset{T \times N}{\boldsymbol{X}} = [\boldsymbol{X}_{\mathcal{S}}, \boldsymbol{X}_{-\mathcal{S}}]$. Finally, since each horizon of the LPs is estimated separately, we suppress the dependence on h in our notation.

2.1 Local Projection Estimation

We start by estimating eq. (4) with the lasso to obtain an initial estimate of $\boldsymbol{\beta}$, while leaving the parameters of interest $\boldsymbol{\beta}_{\mathcal{S}}$ unpenalized. The first stage lasso estimator is defined as

$$\hat{\boldsymbol{\beta}}^{(L)} = [\hat{\boldsymbol{\beta}}_{\mathcal{S}}^{(L)\prime}, \hat{\boldsymbol{\beta}}_{-\mathcal{S}}^{(L)\prime}]' = \underset{\boldsymbol{\beta}^* \in \mathbb{R}^N}{\arg\min} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^*\|_2^2 / T + 2\lambda \|\boldsymbol{W}\boldsymbol{\beta}^*\|_1, \tag{5}$$

where $\underset{N \times N}{\boldsymbol{W}}$ is a diagonal matrix with $\boldsymbol{W}_{i,i} = 0$ for $i \in \mathcal{S}$ and 1 otherwise. The Frish-Waugh-Lovell-style result of Yamada (2017) ensures that the penalized and unpenalized estimates can be respectively obtained by

$$\hat{\boldsymbol{\beta}}_{-\mathcal{S}}^{(L)} = \underset{\boldsymbol{\beta}^* \in \mathbb{R}^{(N-S)}}{\arg\min} \|\boldsymbol{M}_{(\boldsymbol{X}_{\mathcal{S}})}\boldsymbol{y} - \boldsymbol{M}_{(\boldsymbol{X}_{\mathcal{S}})}\boldsymbol{X}_{-\mathcal{S}}\boldsymbol{\beta}^*\|_2^2 / T + 2\lambda \|\boldsymbol{\beta}^*\|_1,$$

$$\hat{\boldsymbol{\beta}}_{\mathcal{S}}^{(L)} = \hat{\boldsymbol{\Sigma}}_{\mathcal{S}}^{-1} \boldsymbol{X}_{\mathcal{S}}'(\boldsymbol{y} - \boldsymbol{X}_{-\mathcal{S}}\hat{\boldsymbol{\beta}}_{-\mathcal{S}}^{(L)})/T,$$

where the residual maker $\boldsymbol{M}_{(\boldsymbol{X}_{\mathcal{S}})} := (I - \boldsymbol{X}_{\mathcal{S}}(\boldsymbol{X}_{\mathcal{S}}'\boldsymbol{X}_{\mathcal{S}})^{-1}\boldsymbol{X}_{\mathcal{S}}')$, and $\hat{\boldsymbol{\Sigma}}_{\mathcal{S}} := \frac{\boldsymbol{X}_{\mathcal{S}}'\boldsymbol{X}_{\mathcal{S}}}{T}$. We can therefore use regular lasso estimation on the transformed response $\boldsymbol{M}_{(\boldsymbol{X}_{\mathcal{S}})}\boldsymbol{y}$ and predictors $\boldsymbol{M}_{(\boldsymbol{X}_{\mathcal{S}})}\boldsymbol{X}_{-\mathcal{S}}$ to obtain $\hat{\boldsymbol{\beta}}_{-\mathcal{S}}^{(L)}$.

We next "desparsify" this estimate to allow for uniformly valid inference. The desparsified lasso estimator is defined as

$$\hat{\boldsymbol{\beta}}_{\mathcal{S}} := \hat{\boldsymbol{\beta}}_{\mathcal{S}}^{(L)} + \hat{\boldsymbol{\Theta}}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{(L)})/T, \tag{6}$$

where $\underset{S \times N}{\hat{\boldsymbol{\Theta}}}$ is a submatrix of the approximate inverse covariance matrix of $\hat{\boldsymbol{\Sigma}} := \frac{\boldsymbol{X}'\boldsymbol{X}}{T}$. We obtain $\hat{\boldsymbol{\Theta}}$ from so-called nodewise regressions, which can be seen as a linear projection of $\boldsymbol{x}_j$ onto $\boldsymbol{X}_{-j}$, where $\boldsymbol{x}_j$ is the jth column of $\boldsymbol{X}$ and $\boldsymbol{X}_{-j}$ is $\boldsymbol{X}$ without the jth column

$$\boldsymbol{x}_j = \boldsymbol{X}_{-j}\boldsymbol{\gamma}_j + \boldsymbol{v}_j, \text{ where } \boldsymbol{\gamma}_j = \underset{\boldsymbol{\gamma}_j^* \in \mathbb{R}^{N-1}}{\arg\min} \mathbb{E}\|\boldsymbol{x}_j - \boldsymbol{X}_{-j}\boldsymbol{\gamma}_j^*\|_2^2 / T. \tag{7}$$

The lasso estimates of the nodewise regressions are given by

$$\hat{\boldsymbol{\gamma}}_j = \underset{\boldsymbol{\gamma}_j^* \in \mathbb{R}^{N-1}}{\arg\min} \|\boldsymbol{x}_j - \boldsymbol{X}_{-j}\boldsymbol{\gamma}_j^*\|_2^2 / T + 2\lambda_j \|\boldsymbol{\gamma}_j^*\|_1,$$

Finally, we construct $\hat{\boldsymbol{\Theta}} := \underset{S \times S}{\hat{\boldsymbol{\Upsilon}}^{-2}} \underset{S \times N}{\hat{\boldsymbol{\Gamma}}}$, where $\hat{\boldsymbol{\Upsilon}}^{-2} := \mathrm{diag}(1/\hat{\tau}_1^2, \dots 1/\hat{\tau}_S^2)$, with $\hat{\tau}_j^2 = \|\hat{\boldsymbol{v}}_j\|_2^2 / T +$

$\lambda_j \left\| \hat{\boldsymbol{\gamma}}_j \right\|_1$, $\hat{\boldsymbol{v}}_j = \boldsymbol{x} - \boldsymbol{X} \hat{\boldsymbol{\gamma}}_j$ and $\hat{\boldsymbol{\Gamma}}$ stacks the $\hat{\boldsymbol{\gamma}}_j$'s row-wise with ones on the left diagonal section

$$\hat{\boldsymbol{\Gamma}} := \begin{bmatrix} 1 & \dots & -\hat{\gamma}_{1,S} & \dots & -\hat{\gamma}_{1,N} \\ \vdots & \ddots & \vdots & & \vdots \\ -\hat{\gamma}_{S,1} & \dots & 1 & \dots & -\hat{\gamma}_{S,N} \end{bmatrix}.$$

The specific structure of the LPs consisting of the same set of regressors at each horizon h allows for computational and efficiency gains in obtaining the desparsified lasso. The initial lasso in eq. (5) should be obtained for each horizon, but the population linear projection coefficients $\boldsymbol{\gamma}_j$ in the nodewise regression eq. (7) do not change with the horizon. It is therefore sufficient to compute the nodewise lasso estimator once. This can be done best at horizon zero where most time points are available for estimating the nodewise regression. We therefore partly avoid the loss of efficiency at further horizons where the most recent observation at each new horizon would be lost for estimation otherwise. As an example, for the LP in eq. (1), we estimate

$$x_t = \boldsymbol{\gamma}^{(0)\prime} \boldsymbol{x}_{s,t} + \sum_{k=1}^{K} \boldsymbol{\gamma}^{(k)\prime} \boldsymbol{z}_{t-k} + v_{1,t}, \tag{8}$$

then store $\hat{\boldsymbol{\Theta}}$ constructed from $\hat{\boldsymbol{\gamma}}_1 = \left[\hat{\boldsymbol{\gamma}}^{(0)\prime}, \dots, \hat{\boldsymbol{\gamma}}_K^{(K)\prime} \right]'$, and the residual vector $\hat{\boldsymbol{v}}_1$. The former can then be re-used at each horizon in the desparsified lasso estimator in eq. (6), while the latter can be re-used in the estimation of the long-run covariance matrix which enters the asymptotic distribution of the desparsified lasso, see Section 2.2.

Finally, regarding the choice of tuning parameter λ in the initial regression and λ_j in the nodewise regressions, we follow the approach of Adamek et al. (2022b) who propose an iterative plug-in procedure that simulates the quantiles of an appropriate "empirical process" which should be dominated by λ; see their Section 5.1 for details.

2.2 Local Projection Inference

In this section, we discuss the asymptotic properties of the desparsified lasso estimator $\hat{\boldsymbol{\beta}}_{\mathcal{S}}$. The assumptions needed for our analysis are listed in detail in Appendix A, and are discussed briefly below. More detailed discussions can be found in Adamek et al. (2022b), on whose approach our theory is built.

We need assumptions on the DGP (Assumption A.1), the sparsity of the parameter (Assumption A.2), regularity conditions for high-dimensional settings (Assumption A.3) and assumptions on the set of parameters on which inference is conducted (Assumption A.4). Assumption A.1(a) requires that $\boldsymbol{z}_t = \left(\boldsymbol{x}'_{s,t}, x_t, y_t, \boldsymbol{x}'_{f,t} \right)'$ the regressors are contemporaneously uncorrelated with the

7

error term, and that the error terms and the variables in $\boldsymbol{z}_t$ have finite unconditional moments up to a certain order. Crucially, Assumption A.1(a) allows for serial correlation in the error terms, which is a typical feature of LP regressions. Assumption A.1(c) assumes $\boldsymbol{z}_t$ to be near-epoch-dependent (NED). An NED process can be interpreted as a process that is well-approximated by a mixing process, but does not have to be mixing itself. Unlike mixing conditions, which may exclude even very simple autoregressive processes, NED conditions permit very general forms of dependence often encountered in macroeconomic research, see Adamek et al. (2022b) for a more detailed discussion.

Comparing our DGP assumptions to the SVMA assumptions of Example 1, one can verify that the moment condition is satisfied trivially under Gaussianity, and the NED assumption can be shown to hold when the VMA coefficients $\boldsymbol{A}_k$ converge to 0 sufficiently quickly; see Example 17.3 of Davidson (2002). If one would relax the Gaussianity assumption, a trade-off would occur between the number of moments and the allowed dependence – the more moments, the more persistent the process may be.

Second, we impose a sparsity assumption on the parameter $\boldsymbol{\beta}$ (Assumption A.2), as typically required for lasso estimators. We derive our results under weak sparsity, thereby recognizing that the true parameters are likely not exactly zero. In particular, we assume that $\|\boldsymbol{\beta}\|_r^r$ is sufficiently small for some $0 \leq r < 1$. While for $r = 0$ our results boil down to the requirement of exact sparsity– which assumes few non-zero entries in the parameter vector – for larger values of r, we allow for many non-zero entries in $\boldsymbol{\beta}$ provided they are relatively small. The smaller r, the more restrictive this assumption therefore is. We also require weak sparsity in the nodewise regressions, which amounts to assuming that the inverse population covariance matrix $\boldsymbol{\Sigma}^{-1} := (\mathbb{E}\boldsymbol{X}'\boldsymbol{X}/T)^{-1}$ is sparse in terms of (induced) r-norm. Since eq. (1) has a similar regressor structure as a VAR model, this can be shown to hold under conditions similar to the sparse VAR in Example 6 of Adamek et al. (2022b).

Next, we require that the minimum eigenvalue of $\boldsymbol{\Sigma}$ is bounded away from 0 (Assumption A.3). This assumption ensures that the population covariance matrix of the regressors remains well-behaved with increasing N. Again, thanks to the similarity of eq. (1) to a VAR, lower bounds on the minimum eigenvalue can be derived using the results on page 6 of Masini et al. (2022). Finally, we also require that the number of unpenalized parameters of interest to be bounded (Assumption A.4). In LPs, this is trivially satisfied, for instance $S = 1$ in eq. (1), and $S = 3$ in eq. (3).[2]

---

[2]While the impulse responses are the two parameters of interest, we may also want to leave the dummy parameter unpenalized.

Under these assumptions, Theorem 1 establishes the asymptotic normality of our desparsified lasso estimator, which allows for valid asymptotic inference.

Theorem 1. *Under Assumptions A.1 to A.5, for the model in eq. (4), we have*

$$\sup_{\substack{\boldsymbol{\beta}\in\boldsymbol{B}_N(r,s_r) \\ z\in\mathbb{R}}} \left| \mathbb{P}\left(\sqrt{T}\frac{\boldsymbol{r}'(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S)}{\sqrt{\boldsymbol{r}'\hat{\boldsymbol{\Upsilon}}^{-2}\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{\Upsilon}}^{-2}\boldsymbol{r}}} \le z \right) - \Phi(z) \right| = o_p(1),$$

where $\Phi(\cdot)$ is the CDF of $N(0,1)$, $\boldsymbol{r}\in\mathbb{R}^S$ is chosen to test hypotheses of the form $\boldsymbol{r}'\boldsymbol{\beta}_S = 0$, and $\hat{\boldsymbol{\Omega}}$ is a consistent long-run covariance estimator.

As in Adamek et al. (2022b), we use the Newey-West long run covariance estimator

$$\underset{S\times S}{\hat{\boldsymbol{\Omega}}} := \sum_{l=1-Q_T}^{Q_T-1} (1 - l/Q_T)\,\hat{\tilde{\boldsymbol{\Xi}}}(l),$$

for some $Q_T < T$, where $\hat{\tilde{\boldsymbol{\Xi}}}(l) := \frac{1}{T-l}\sum_{t=l+1}^{T}\hat{\boldsymbol{w}}_t\hat{\boldsymbol{w}}'_{t-l}$, $\hat{\boldsymbol{w}}_t := (\hat{v}_{1,t},\dots,\hat{v}_{S,t})'\,\hat{u}_t$. To obtain our asymptotic results, we generally require that the bandwidth Q_T grows as $T\to\infty$ at a sufficiently slow rate, see Assumption A.5 in Appendix A for details. Regarding the choice of Q_T in finite samples, we use the bandwidth estimator of Andrews (1991).

For the LP in eq. (1), Theorem 1 permits asymptotic confidence intervals of the form

$$CI_h(\alpha) := \left[\hat{\phi}_h \pm z_{\alpha/2}\sqrt{\frac{\hat{\omega}_h/\hat{\tau}_h^4}{T}}\,\right], \tag{9}$$

where $\hat{\omega}_h$ and $\hat{\tau}_h^{-2}$ are the scalar versions of $\hat{\boldsymbol{\Omega}}_h$ and $\hat{\boldsymbol{\Upsilon}}_h^{-2}$ respectively, and $z_{\alpha/2} := \Phi^{-1}(1-\alpha/2)$. Note the dependence on h in this result, since the residuals $\hat{\boldsymbol{u}} = \hat{\boldsymbol{u}}_h$ are different at each horizon, and therefore $\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\Omega}}_h$ and $\hat{\boldsymbol{\Upsilon}}^{-2} = \hat{\boldsymbol{\Upsilon}}_h^{-2}$ as well.

3 Simulations

We perform a simulation study to analyse the finite sample performance of the desparsified lasso estimator. Consider the following DGP

$$\begin{pmatrix} y_t \\ \boldsymbol{x}_{s,t} \end{pmatrix} = \underset{P\times 1}{\boldsymbol{x}_t} = \sum_{k=1}^{4}\boldsymbol{A}_k\boldsymbol{x}_{t-k} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{iid}{\sim} N(\boldsymbol{0},\boldsymbol{I}), \tag{10}$$

where the autoregressive parameter matrices $\boldsymbol{A}_k$ are tapered Toeplitz matrices for which we consider two different settings. In the first setting, we take $(\boldsymbol{A}_k)_{i,j} = \rho_k^{|i-j|+1}$ if $|i-j| < P/2$ and 0 otherwise, and $(\rho_1,\rho_2,\rho_3,\rho_4) = (0.2, 0.15, 0.1, 0.05)$. In the second setting, we simply switch the signs of all

entries in the second and fourth autoregressive parameter matrices. We consider different values for the number of variables $P = \{20, 40, 100\}$ and time series length $T = \{100, 200, 500\}$.[3]

Equation (10) can be seen as a simple structural VAR with structural shocks $\boldsymbol{\epsilon}_t$. The true impulse responses are obtained by inverting the lag polynomial, rewriting it as $\boldsymbol{x}_t = \sum_{k=0}^{\infty} \boldsymbol{B}_k \boldsymbol{\epsilon}_{t-k} = \boldsymbol{B}(L)\boldsymbol{\epsilon}_t$, where $\boldsymbol{B}(L) = \boldsymbol{A}(L)^{-1} = \left(\boldsymbol{I} - \sum_{k=1}^{4} \boldsymbol{A}_k L^k \right)^{-1}$ and $\boldsymbol{B}_0 = \boldsymbol{I}$. We analyse the response of $y_t = x_{1,t}$ to the first shock $\epsilon_{1,t}$ using local projections

$$ y_{t+h} = \phi_h y_t + \boldsymbol{x}'_{s,t} \boldsymbol{\eta}_h + \sum_{k=1}^{K} (y_{t-k}, \boldsymbol{x}'_{s,t-k}) \boldsymbol{\delta}_{h,k} + u_{h,t}, \quad h = 1, \dots, h_{\max}, $$

and obtain estimates of the impulse responses $\hat{\phi}_1, \dots, \hat{\phi}_{h_{\max}}$ with $h_{\max} = 10$ and $K = 4$.[4] We investigate whether the 95% point-wise confidence intervals corresponding to the impulse response parameters at horizons 0 to 10 cover their true values at the correct rates. Results are reported for 1000 replications. Figure 1 presents the coverage rates for the proposed desparsified lasso (blue), which leaves the impulse response parameters of interest unpenalized, in comparison to the standard desparsified lasso (red). Figure C.1 in Appendix C includes the mean interval widths.

The proposed desparsified lasso estimator attains good coverage rates close to the nominal level of 95% for all combinations of P and T. For both DGPs, coverage rates are usually the lowest around 70% to 90% for horizon 1, but revert to the nominal level for larger horizons. Good coverage at further horizons is to be expected, since the local projection coefficients will tend to become more sparse in stationary models. Importantly, considerable gains are obtained compared to the standard desparsified lasso. Especially at shorter horizons – which are usually of most interest in practice – coverage rates of the latter suffer with drops to 25% in the first DGP or even more severe ones in the second DGP. We also see in Figure C.1 of Appendix C that the interval widths are broadly similar or even smaller than the ones for the standard desparsified lasso, meaning we obtain better coverage without losing power.

The poor coverage rates of the standard desparsified lasso occur since the parameters of interest are not selected in the initial lasso regression. A simple alteration of the desparsified lasso that leaves this parameter unpenalized thus brings the coverage rates much closer to the nominal level. Note that the standard desparsified lasso has coverage exceeding our proposed estimator at further horizons; this is because the true impulse response becomes close to zero, where the bias towards zero of the lasso is beneficial. This does not run counter to our conclusion, as we believe more uniform coverage over different parameter values is desirable.

---

[3]The number of regressors in the model is $N = 5 \times P$, due to the included lags.

[4]At horizon 0 the response is $(\boldsymbol{B}_0)_{1,1} = 1$ by our identification scheme, and thus does not need to be estimated.
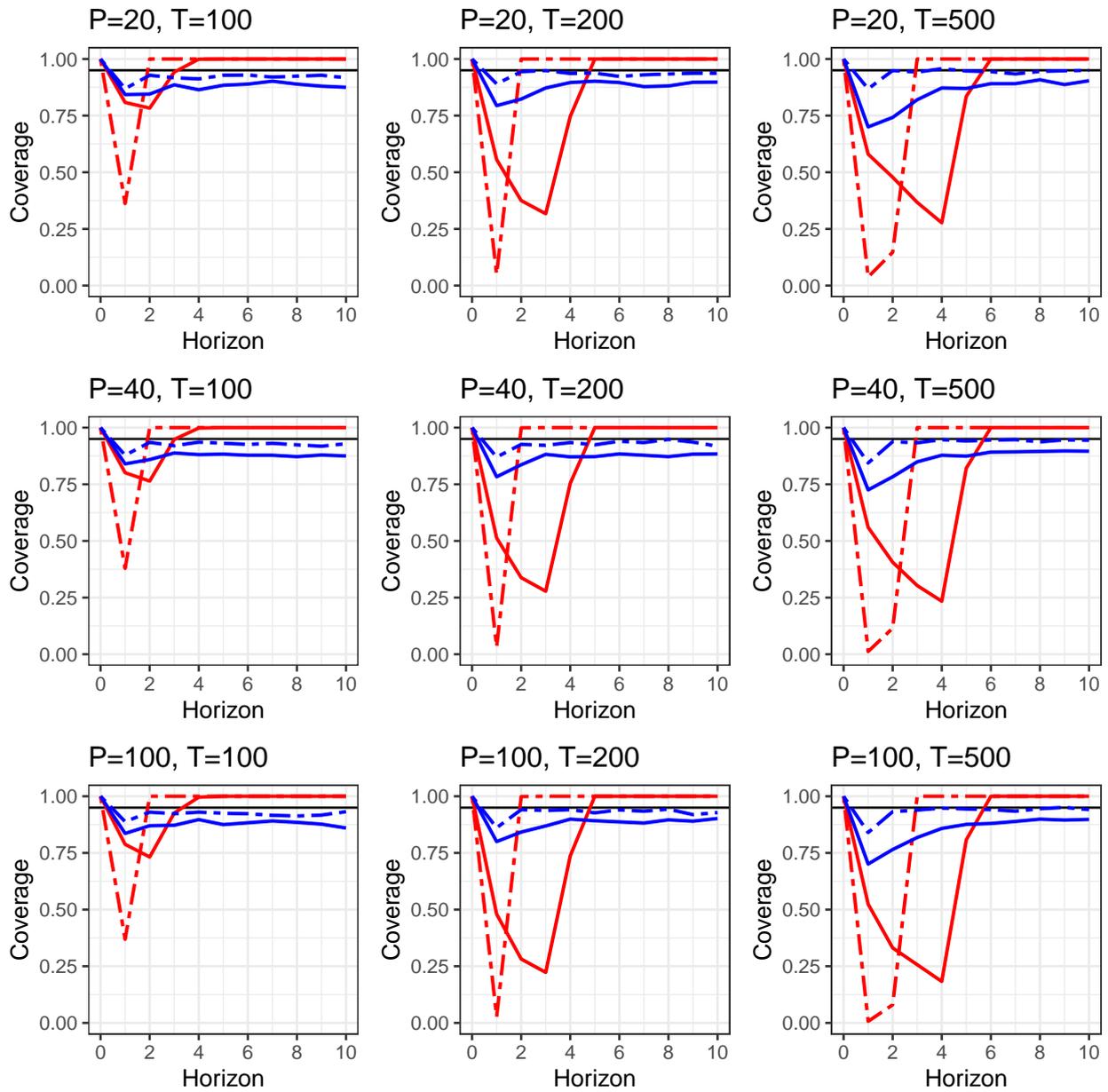
Figure 1: Coverage rates of the standard desparsified lasso (red) and the proposed desparsified lasso with ϕ_h unpenalized (blue). Dashed lines indicate results for the second sign-switching DGP.

4 Structural Impulse Responses Estimated by HDLPs

We apply the desparsified lasso for HDLPs to two canonical macroeconomic applications. In Section 4.1, we build on the work of Bernanke et al. (2005) on monetary policy analysis, in Section 4.2 we build on the work by Ramey and Zubairy (2018) on government spending. All analyses are performed in R (R Core Team, 2022) using the package `desla` (Adamek et al., 2022a).

4.1 Impulse Responses to a Shock in Monetary Policy

We consider the work of Bernanke et al. (2005), wherein they estimate impulse responses to a monetary policy shock identified by the federal funds rate in a factor-augmented SVAR. In this paper, we use HDLPs to perform the impulse response analysis and expand the dataset used in the original paper to the more recent and larger FRED-MD database (McCracken and Ng, 2016), thereby considering monthly data from January 1960 to December 2019 on 122 macroeconomic variables. For a full description of the data, see Appendix D.

We estimate the HDLP in eq. (1), where as response variables y_t we take the federal funds rate (FFR), industrial production (IP) and the consumer price index (CPI). As shock variable of interest x_t we consider the federal funds rate (FFR). Furthermore, we use the identification strategy of Bernanke et al. (2005), which consists of classifying variables as "slow" in $\boldsymbol{x}_{s,t}$ or "fast" in $\boldsymbol{x}_{f,t}$, according to whether these variables can respond within the same month to an unexpected change in the FFR.[5] For variables in FRED-MD that can be directly matched to their counterparts in the DRI/McGraw Hill Basic Economics Database, we use the same slow/fast classification as Bernanke et al. (2005). For new variables in FRED-MD, we apply the same general classification rule; i.e., variables in the categories Prices, Output & Income, Labour Market, and Consumption are classified as slow, whereas those in the categories Interest & Exchange Rates, Money & Credit, Stock Market, and Housing as fast.[6] We apply the transformation codes of Bernanke et al. (2005) and McCracken and Ng (2016) (for new variables) to remove stochastic trends. Using $K = 13$ lags. the resulting HDLP consists of $N = 1654$ regressors, while the time series length is $T = 707$.[7]

We additionally extend the analysis to a nonlinear state-dependent HDLP, inspired by recent work concerning the effectiveness of the FFR as a monetary policy instrument at the zero lower

---

[5]As discussed before, this strategy is equivalent to recursive (partial) identification in an SVAR, as shown in Example 1 of Plagborg-Møller and Wolf (2021). In fact, this application can also be seen as a high-dimensional extension of Christiano et al. (2005).

[6]One exception in the Category Prices is the series "OILPRICEx", which we classify as fast.

[7]The exact number of variables in the model depends on which variable is taken as the response, and the effective number of observations depends on the horizon. These numbers are for the model where FFR is the response, at horizon 1.

bound (ZLB) (Gambacorta et al., 2014; Inoue and Rossi, 2021; Wu and Xia, 2016). Specifically, we estimate the model in Example 2, taking as I_{t-1} the indicator function when the lag of FFR is at or below 0.25%, thereby following the choice of Wu and Xia (2016). As such, we obtain different impulse response estimates $\hat{\phi}_{A,h}$ and $\hat{\phi}_{B,h}$ depending on the state of the economy. In this state-dependent HDLP, we effectively double the number of regressors compared to the original HDLP, with $N = 3309$.

We compare the HDLP impulse responses obtained from the desparsified lasso to the ones obtained from a 3-factor FAVAR as used in Bernanke et al. (2005). Additional details about the FAVAR estimation are provided in Appendix E. Figure 2 shows the impulse responses of FFR, IP, and CPI to a shock in the FFR of a size such that the FFR has unit response at horizon 0. Since we use the first difference of IP and CPI in the model, we cumulate the impulse response which then corresponds to the response in the original (level) variables.[8] The comparable figure in Bernanke et al. (2005) is Figure 1.

The impulse responses from the FAVAR (bottom panel Figure 2) and linear HDLP (black line in top panel) are generally similar. The impulse response of the FFR peaks at horizon 1 and then steadily declines to zero, thereby being in line with Figure 1 in Bernanke et al. (2005). For the HDLP, the response stays at a higher level for a longer period than for the FAVAR. Next, the response of IP is also in line with Bernanke et al. (2005), with the largest drop in response around horizon 20, before eventually returning to zero. Notably, the response obtained with the HDLP specification is considerably smaller than the one for the FAVAR. Finally for CPI, in contrast to the results of Bernanke et al. (2005), we see a pronounced price puzzle for both the FAVAR and HDLP specifications, though the HDLP response is smaller in magnitude and turns negative around horizon 20.

When examining the impulse responses in the state-dependent model, we see considerably different responses between the states. Above the ZLB, responses are similar to the linear HDLP specification, while the responses are much larger in magnitude and more uncertain at the ZLB. As we only have 85 time points at the ZLB in our sample and very little variation in the FFR, it is not surprising to see much more erratic estimates during this state. The response of the FFR is not significantly different from zero past horizon 0, and the response of IP becomes large and positive after horizon 15, unlike the linear model which estimates a negative response at the same

---

[8]Cumulative impulse response functions for the HDLP are obtained by cumulating the dependent variable, i.e. taking $\sum_{\ell=0}^{h} y_{t+\ell}$ on the LHS of eq. (1). For the FAVAR, we take the cumulative sum of the regular impulse responses from previous horizons. To ensure that these impulse responses are comparable, we scale the FAVAR impulse responses by the response of the FFR at horizon 0.

Linear model & Above ZLB

At ZLB
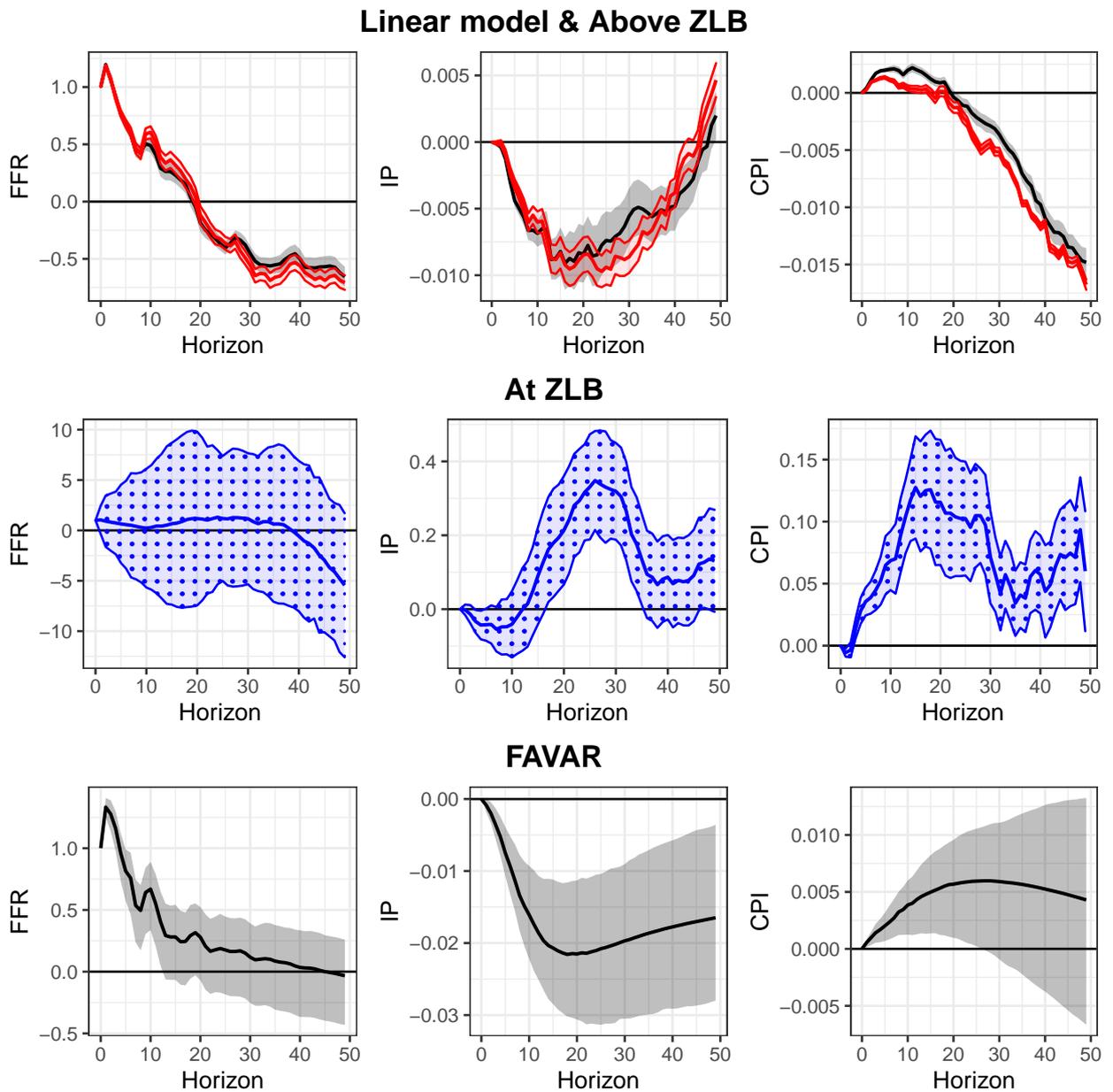
FAVAR



Figure 2: Estimated impulse responses with 95% confidence intervals of FFR, IP, and CPI to an identified monetary policy shock by the FFR for the HDLP and FAVAR specifications. Black impulse responses with grey shaded confidence intervals correspond to the linear model, blue with dotted confidence intervals to the ZLB state, red with solid-bordered confidence intervals to the above ZLB state.

horizon. Regarding CPI, we see a reduced price puzzle above the ZLB compared to the linear HDLP specification, and correspondingly a pronounced price puzzle at the ZLB. We find these results to be in line with the idea that the FFR becomes an ineffective monetary policy instrument at or near the ZLB.

4.2 Impulse Responses to a Shock in Government Spending

We next build on the work of Ramey and Zubairy (2018), who estimate impulse responses to a shock in US government spending identified based on military spending news. In this paper, we first augment the authors' main LP specification with more lags as a robustness check, before considering an extended state-dependent HDLP specification with interacting states related to unemployment and recession conditions. We use the quarterly data provided by the authors at econweb.ucsd.edu/∼vramey which covers the period 1889Q1 to 2015Q4.

We estimate the non-linear state-dependent HDLP in Example 2 for different responses y_t, namely real per capita GDP and government spending, while the shock variable of interest x_t is the military spending news shock. We include taxes as controls in $\boldsymbol{x}_{f,t}$, while no variables are included in $\boldsymbol{x}_{s,t}$.[9] The state dummy variable I_t distinguishes between low and high unemployment states, with $I_t = 1$ when unemployment in period t is larger than 6.5%. For full details on the construction of the shock variable, and the treatment of the other variables, see section II.B of Ramey and Zubairy (2018). We use $K = 40$ lags, such that the resulting HDLP consists of $N = 464$ regressors, while the time series length is $T = 161$.[10]

Figure 3 shows the impulse responses to a shock in government spending. Note that the military spending news variable is scaled by GDP such that the impulse responses can be seen as a reaction to a shock of size 1% of GDP. The comparable figure in Ramey and Zubairy (2018) is Figure 5. The general shape and magnitude of these impulse responses are similar to the LP model with only 4 lags used in Ramey and Zubairy (2018). In our analysis, the peak of the responses in the high unemployment state occurs 2-3 quarters earlier, and experiences a sharper drop after horizon 15. We also find that the impulse response is not significant at horizon 0 in the linear and low-unemployment state.

Next to allowing the inclusion of more lags, the HDLP framework can easily be extended to

---

[9]While Ramey and Zubairy (2018) do not include taxes in their main analysis, they mention in footnote 11 that their results were robust to their inclusion. We prefer to include them as additional controls, following the VAR of Blanchard and Perotti (2002).

[10]As in Section 4.1, these correspond to the model estimated at horizon 1.
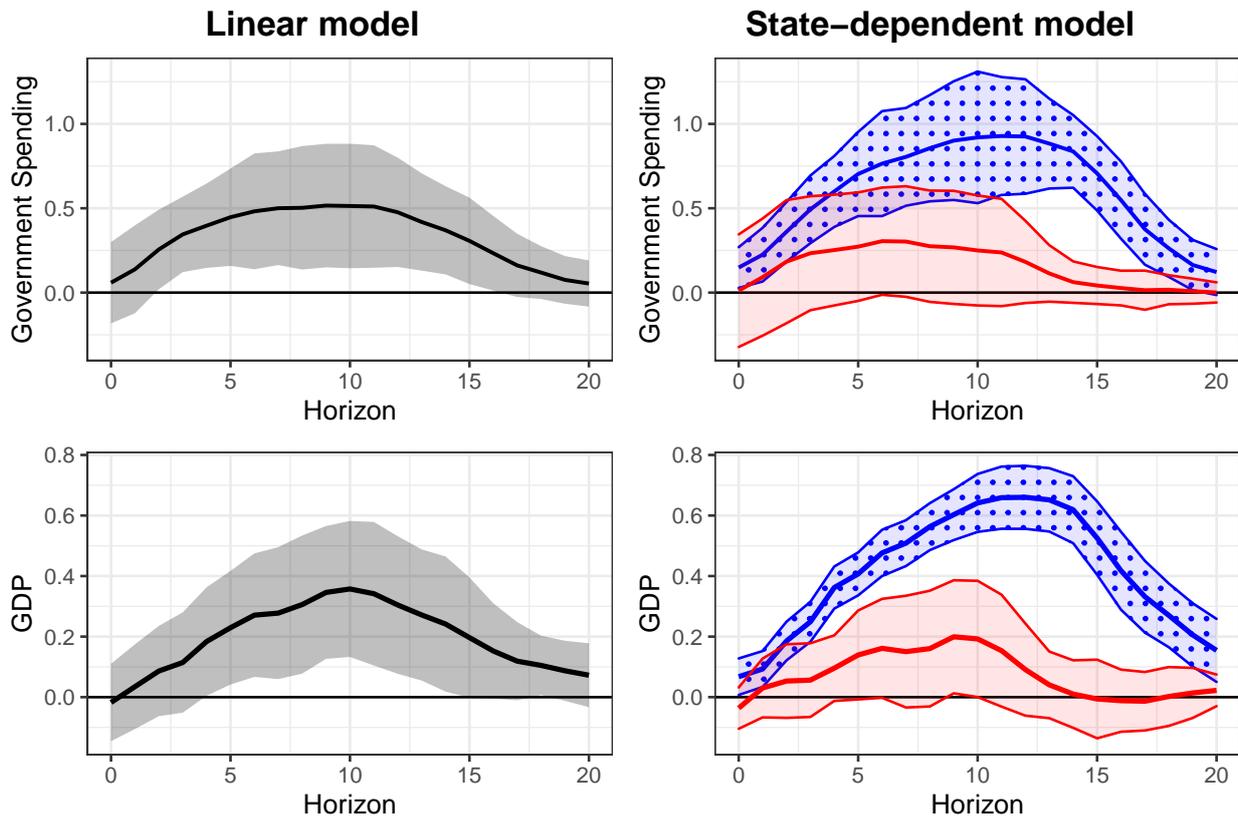
Figure 3: Estimated impulse responses with 95% confidence intervals of government spending and GDP to a government spending shock, in a model with 40 lags. The left column shows the linear HDLP specification (black), and the right column compares the high-unemployment state (blue with dotted intervals) and the low-unemployment state (red with solid-bordered intervals).

allow for S states. The LP is then given by

$$y_{t+h} = \sum_{i=1}^{S} \alpha_i I_{i,t-1} + \sum_{i=1}^{S} I_{i,t-1} \left[\phi_{i,h} x_t + \sum_{k=1}^{K} \boldsymbol{\delta}'_{i,h,k} \boldsymbol{z}_{t-k} \right] + u_{h,t}.$$

In addition to the unemployment state dummy $I_t^{(U)}$, we consider recession dummy $I_t^{(R)}$.[11] Letting them interact results in $S = 4$ distinct state dummies:

$$I_{1,t} = I_t^{(U)} I_t^{(R)}, \ I_{2,t} = I_t^{(U)}(1 - I_t^{(R)}), \ I_{3,t} = (1 - I_t^{(U)}) I_t^{(R)}, \ I_{4,t} = (1 - I_t^{(U)})(1 - I_t^{(R)}).$$

Figure 4 displays the impulse response functions for the non-linear HDLP with four states. The impulse responses out of recessions resemble those from the state-dependent model in Figure 3, both in shape and magnitude, but we see a different pattern in the recession state. During a recession with low unemployment, there is a negative impact response for both government spending and GDP, which becomes insignificant after a few periods. In the high unemployment state, in contrast, we see a large, positive, and persistent effect on both variables.

---

[11]As in Ramey and Zubairy (2018), we follow the NBER classification of recession periods.
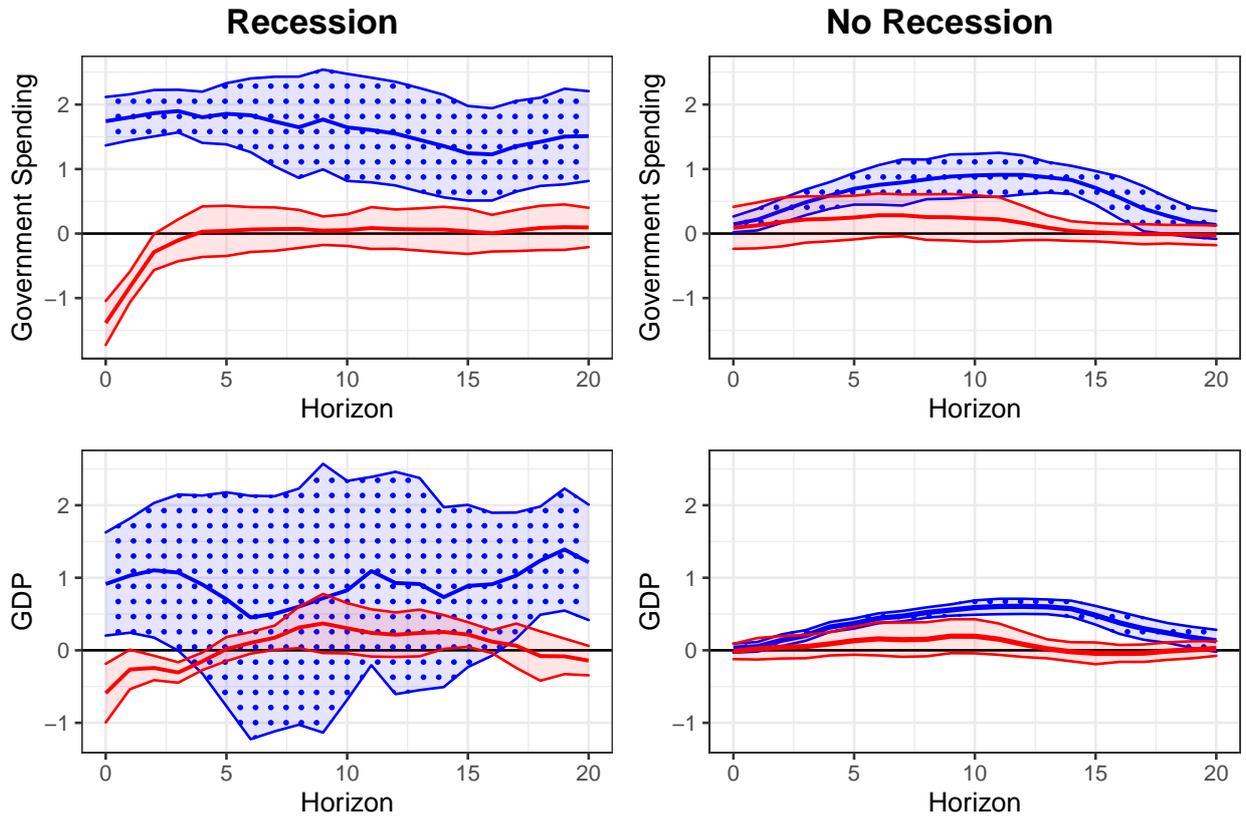
Figure 4: Estimated Impulse responses with 95% confidence intervals of government spending and GDP to a government spending shock. Blue curves with dotted intervals indicate the impulse response in the high unemployment state, red curves with solid-bordered intervals in the low unemployment state.

5 Conclusion

In this paper, we propose a modified version of the desparsified lasso to estimate local projections in high dimensions. The modification is simple, namely leaving the (small number of) dynamic impulse response parameter(s) of interest unpenalized in the initial lasso regression, yet provides considerable improvements in finite sample performance in terms of better coverage. The modified desparsified lasso estimator still comes with desirable asymptotic normality and uniformly valid asymptotic inference. Finally, we show how this method performs on canonical macroeconomic applications where we use the high-dimensional local projections framework to estimate structural impulse responses.

Acknowledgements

tricht Workshop on Dimensionality Reduction and Inference in High-Dimensional Time Series. We gratefully acknowledge the comments by participants at these seminars and conferences. In addition, we thank Otilia Boldea and Lenard Lieb for helpful discussions. All remaining errors are our own.

References

Adamek, R., S. Smeekes, and I. Wilms (2022a). *desla: Desparsified Lasso Inference for Time Series*. R package version 0.2.0.

Adamek, R., S. Smeekes, and I. Wilms (2022b). Lasso inference for high-dimensional time series. *Journal of Econometrics*, Forthcoming.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Angrist, J. D., Òscar Jordà, and G. M. Kuersteiner (2018). Semiparametric estimates of monetary policy effects: string theory revisited. *Journal of Business & Economic Statistics 36*, 371–387.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*, 71–92.

Basu, S. and G. Michailidis (2015). Regularized estimation in sparse high-dimensional time series models. *Annals of Statistics 43*(4), 1535–1567.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies 81*, 608–650.

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

Bernanke, B. S. and I. Mihov (1998). Measuring monetary policy. *Quarterly Journal of Economics 113*, 869–902.

Blanchard, O. and R. Perotti (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics 117*, 1329–1368.

Bühlmann, P. and S. van De Geer (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer.

Chan, J. C. (2020). Large Bayesian vector autoregressions. In P. Fuleky (Ed.), *Macroeconomic Forecasting in the Era of Big Data*, Volume 52 of *Advanced Studies in Theoretical and Applied Econometrics*, Chapter 4, pp. 95–125. Springer.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy 113*, 1–45.

Chudik, A. and M. H. Pesaran (2016). Theory and practice of GVAR modelling. *Journal of Economic Surveys 30*, 165–197.

Davidson, J. (2002). *Stochastic Limit Theory* (2nd ed.). Oxford: Oxford University Press.

Forni, M. and L. Gambetti (2014). Sufficient information in structural VARs. *Journal of Monetary Economics 66*, 124–136.

Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory 25*, 1319–1347.

Gambacorta, L., B. Hofmann, and G. Peersman (2014). The effectiveness of unconventional monetary policy at the zero lower bound: a cross-country analysis. *Journal of Money, Credit and Banking 46*, 615–642.

Inoue, A. and B. Rossi (2021). A new approach to measuring economic policy shocks, with an application to conventional and unconventional monetary policy. *Quantitative Economics 12*, 1085–1138.

Javanmard, A. and A. Montanari (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research 15*, 2869–2909.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review 95*(1), 161–182.

Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Themes in Modern Econometrics. Cambridge University Press.

Kock, A. B. and L. Callot (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics 186*, 325–344.

Kock, A. B., M. Medeiros, and V. G (2020). Penalized regressions. In P. Fuleky (Ed.), *Macroeconomic Forecasting in the Era of Big Data*, Volume 52 of *Advanced Studies in Theoretical and Applied Econometrics*, Chapter 7, pp. 193–228. Springer.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multi-variate models. *Journal of Econometrics 74*, 119–147.

Krampe, J., E. Paparoditis, and C. Trenkler (2022). Structural inference in sparse high-dimensional vector autoregressions. *Journal of Econometrics*, Forthcoming.

Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometric Theory 21*, 21–59.

Masini, R. P., M. C. Medeiros, and E. F. Mendes (2022). Regularized estimation of high-dimensional vector autoregressions with weakly dependent innovations. *Journal of Time Series Analysis 43*, 532–557.

McCracken, M. W. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*, 574–589.

Montiel Olea, J. L. and M. Plagborg-Møller (2021). Local projection inference is simpler and more robust than you think. *Econometrica 89*, 1789–1823.

Plagborg-Møller, M. and C. K. Wolf (2021). Local projections and VARs estimate the same impulse responses. *Econometrica 89*, 955–980.

R Core Team (2022). R: *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Volume 2, pp. 71–162. Elsevier.

Ramey, V. A. and S. Zubairy (2018). Government spending multipliers in good times and in bad: evidence from US historical data. *Journal of Political Economy 126*, 850–901.

Romer, C. D. and D. H. Romer (2004). A new measure of monetary shocks: derivation and implications. *American Economic Review 94*(4), 1055–1084.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*, 1–48.

Sims, C. A. and T. Zha (2006). Were there regime switches in U.S. monetary policy? *American Economic Review 96*(1), 54–81.

Stock, J. and M. Watson (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Volume 2, pp. 415–525. Elsevier.

Stock, J. H. and M. W. Watson (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *Economic Journal 128*, 917–948.

van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Annals of Statistics 42*, 1166–1202.

van de Geer, S. A. (2016). *Estimation and testing under sparsity.* Springer.

Wu, J. C. and F. D. Xia (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking 48*, 253–291.

Yamada, H. (2017). The Frisch–Waugh–Lovell theorem for the lasso and the ridge regression. *Communications in Statistics-Theory and Methods 46*, 10897–10902.

Appendix A   Assumptions

In the appendix, we use the following additional notation. For a matrix $\boldsymbol{A}$, we let $\|\boldsymbol{A}\|_r = \max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|_r$ for any $r \in [0, \infty]$ and $\|\boldsymbol{A}\|_{\max} = \max_{i,j} |a_{i,j}|$. $\Lambda_{\min}(\boldsymbol{A})$ denotes the minimum eigenvalue of $\boldsymbol{A}$. We frequently make use of arbitrary positive finite constants C (or its sub-indexed version C_i) whose values may change from line to line throughout the paper, but they are always independent of the time and cross-sectional dimension. We say a sequence η_T is of size $-x$ if $\eta_T = O\left(T^{-x-\varepsilon}\right)$ for some $\varepsilon > 0$.

Assumption A.1. There exist some constants $\bar{m} > m > 2$, and $d \geq \max\{1, (\bar{m}/m - 1)/(\bar{m} - 2)\}$ such that

(a) $(\boldsymbol{x}_t', u_t)'$ is a mean zero process with $\mathbb{E}\boldsymbol{x}_t u_t = \boldsymbol{0}$ and
$$\max_{1 \leq j \leq N} \mathbb{E}\left|x_{j,t}\right|^{2\bar{m}} \leq C, \ \mathbb{E}\left|u_t\right|^{2\bar{m}} \leq C,$$

(b) $\max_{1 \leq j \leq S} \mathbb{E}\left|v_{j,t}\right|^{2\bar{m}} \leq C,$

(c) Let $\{\boldsymbol{\epsilon}_{T,t}\}$ denote a $k(T)$-dimensional triangular array that is α-mixing of size $-d/(1/m - 1/\bar{m})$ with σ-field $\mathcal{F}_t^{\boldsymbol{\epsilon}} := \sigma\left\{\boldsymbol{\epsilon}_{T,t}, \boldsymbol{\epsilon}_{T,t-1}, \ldots\right\}$ such that $(\boldsymbol{x}_t', u_t)'$ is $\mathcal{F}_t^{\boldsymbol{\epsilon}}$-measurable. The processes u_t and $x_{j,t}$ are L_{2m}-near-epoch-dependent (NED) of size $-d$ on $\boldsymbol{\epsilon}_{T,t}$ with positive bounded constants, uniformly over $j = 1, \ldots, N$.

Assumption A.2. For some $0 \leq r < 1$ and sparsity level s_r, define the N-dimensional sparse compact parameter space
$$\boldsymbol{B}_N(r, s_r) := \left\{\boldsymbol{\beta} \in \mathbb{R}^N : \|\boldsymbol{\beta}\|_r^r \leq s_r, \ \|\boldsymbol{\beta}\|_{\max} \leq C, \ \exists C < \infty\right\}.$$

Then (a) $\boldsymbol{\beta} \in \boldsymbol{B}_N(r, s_r)$ and (b) $\boldsymbol{\gamma}_j \in \boldsymbol{B}_{N-1}(r, s_{r,j})$ for all $j \in \mathcal{S}$.

Assumption A.3. Let $\Lambda_{\min}(\boldsymbol{\Sigma})$ denote the smallest eigenvalue of $\boldsymbol{\Sigma} = \mathbb{E}\boldsymbol{X}'\boldsymbol{X}/T$, and $\Lambda_{\min}(\boldsymbol{\Omega}_{N,T})$ the smallest eigenvalue of $\boldsymbol{\Omega}_T := \mathbb{E}\left[\frac{1}{T}\left(\sum_{t=1}^{T} \boldsymbol{w}_t\right)\left(\sum_{t=1}^{T} \boldsymbol{w}_t'\right)\right]$ Assume that $1/C \leq \Lambda_{\min}(\boldsymbol{\Sigma}) \leq C$, and $1/C \leq \Lambda_{\min}(\boldsymbol{\Omega}_{N,T}) \leq C$.

Assumption A.4. $S \leq C$, where S is the cardinality of $\mathcal{S}$.

Assumption A.5. Define $\lambda_{\min} := \min\{\lambda, \min_j \lambda_j\}$, $\lambda_{\max} := \max\{\lambda, \max_j \lambda_j\}$, and $s_{r,\max} := \max\{s_r, \max_j s_{r,j}\}$.

Then $Q_T T^{-\frac{1}{2/d+2m/(m-2)}} \to 0$ for some $Q_T \to \infty$, $\lambda \sim \lambda_{\max} \sim \lambda_{\min}$, and

$$0 < r < 1: \quad (\ln\ln T)^{-1} s_{r,\max}^{1/r} \left[\frac{N^{\left(\frac{2}{d}+\frac{2}{m-1}\right)}}{\sqrt{T}} \right]^{\frac{1}{r\left(\frac{1}{d}+\frac{m}{m-1}\right)}} \leq \lambda$$

$$\leq \ln\ln T \left[Q_T^2 \sqrt{T} s_{r,\max} \right]^{-1/(2-r)},$$

$$r = 0: \quad (\ln\ln T)^{-1} \frac{N^{1/m}}{\sqrt{T}} \leq \lambda \leq \ln\ln T \left[Q_T^2 \sqrt{T} s_{0,\max} \right]^{-1/2}.$$

These bounds are feasible when $Q_T^r s_{r,\max} N^{(2-r)\left(\frac{d+m-1}{dm+m-1}\right)} T^{\frac{1}{4}\left(r-\frac{d(m-1)(2-r)}{dm+m-1}\right)} \to 0$, and additionally when $Q_T^2 s_{0,\max} \frac{N^{2/m}}{\sqrt{T}} \to 0$ if $r = 0$.

Appendix B Proofs

Lemma B.1. *Take a vector $\left\{ z \in \mathbb{R}^N : \|z_{S^c}\|_1 \leq 3 \|z_S\|_1 \right\}$, an index set S with cardinality $|S|$ and define the subscript operator $z_S := \left\{ z^* \in \mathbb{R}^N : z_j^* = z_j \mathbb{1}_{\{j \in S\}} \right\}$. Under Assumption A.3, on the set $\mathcal{CC}_T(S) := \left\{ \left\| \hat{\Sigma} - \Sigma \right\|_{\max} \leq C/|S| \right\}$,*

$$\|z_S\|_1 \leq \sqrt{\frac{2|S| z' \hat{\Sigma} z}{\Lambda_{\min}}}.$$

Proof: By Lemma 6.17 and Corollary 6.8 in Bühlmann and van De Geer (2011), this result holds for the "Σ-compatibility condition". By Lemma 6.23 in Bühlmann and van De Geer (2011), this compatibility condition holds under Assumption A.3, and the result follows from the fact that $\Lambda_{\min}$ bounds the compatibility constant from below. $\qquad\square$

Lemma B.2. *Let P be an index set with cardinality $|P|$ such that $\mathcal{S} \subseteq P$. Define the set $\mathcal{E}_T(z) = \left\{ \max_{j \leq N, s \leq T} \left[\left| \sum_{t=1}^{s} u_t x_{j,t} \right| \right] \leq z \right\}$. Under Assumption A.3, on $\mathcal{E}_T(T\frac{\lambda}{4}) \bigcap \mathcal{CC}_T(P)$:*

$$\frac{\|X(\hat{\beta}^{(L)} - \beta)\|_2^2}{T} + \frac{\lambda}{4}\|\hat{\beta}^{(L)} - \beta\|_1 \leq C_1 \lambda^2 |P| + C_2 \lambda \|\beta_{P^c}\|_1.$$

Proof: The proof closely follows the proof of Lemma A.6 in Adamek et al. (2022b), based on Theorem 2.2 of van de Geer (2016), with small modifications to handle the penalty matrix W. From the Lasso optimization problem in eq. (5), we have the Karush-Kuhn-Tucker (KKT) conditions $\frac{X'(y-X\hat{\beta}^{(L)})}{T} = W\hat{\kappa}^* \lambda$, where $\hat{\kappa}^* = (0_{1 \times H}, \hat{\kappa}')'$, and $\hat{\kappa}$ is the subdifferential of $\|\hat{\beta}_{-\mathcal{S}}^{(L)}\|_1$. This then leads to the inequality

$$\frac{(\beta - \hat{\beta}^{(L)})' X'(y - X\hat{\beta}^{(L)})}{T} = (\beta - \hat{\beta}^{(L)})' W \hat{\kappa}^* \lambda = \beta' W \hat{\kappa}^* \lambda - \hat{\beta}^{(L)'} W \hat{\kappa}^* \lambda$$

$$\leq \lambda \|W'\beta\|_1 - \lambda \|W'\hat{\beta}^{(L)}\|_1.$$

To deal with the matrix $\boldsymbol{W}$ in these expressions, we use the bound $\left\|\boldsymbol{W}'(\hat{\boldsymbol{\beta}}_S^{(L)} - \boldsymbol{\beta}_S)\right\|_1 \leq \left\|\hat{\boldsymbol{\beta}}_S^{(L)} - \boldsymbol{\beta}_S\right\|_1$ and the property that $\|\boldsymbol{W}'\boldsymbol{z}_{S^c}\|_1 = \|\boldsymbol{z}_{S^c}\|_1$ for $\boldsymbol{z} \in \mathbb{R}^N$, since $\mathcal{S} \subseteq S$, and those elements set to zero by $\boldsymbol{W}$ are already zero due to subscript S^c operator. Using these additional arguments, we can show that

$$\frac{\left\|\boldsymbol{X}(\hat{\boldsymbol{\beta}}^{(L)} - \boldsymbol{\beta})\right\|_2^2}{T} \leq \frac{5\lambda}{4} \left\|\hat{\boldsymbol{\beta}}_S^{(L)} - \boldsymbol{\beta}_S\right\|_1 - \frac{3\lambda}{4} \left\|\hat{\boldsymbol{\beta}}_{S^c}^{(L)} - \boldsymbol{\beta}_{S^c}\right\|_1 + 2\lambda \left\|\boldsymbol{\beta}_{S^c}\right\|_1,$$

and the remainder of the proof is analogous to the proof of Lemma A.6 in Adamek et al. (2022b), using Lemma B.1 instead of Lemma A.5 in Adamek et al. (2022b). $\qquad\square$

Lemma B.3. *Under Assumptions A.1 to A.4, for any*

$$
\begin{aligned}
0 < r < 1: \quad & \lambda \geq C \ln(\ln(T))^{\frac{d+m-1}{r(dm+m-1)}} \left[s_r \left(\frac{N^{\left(\frac{2}{d}+\frac{2}{m-1}\right)}}{\sqrt{T}} \right)^{\frac{1}{\left(\frac{1}{d}+\frac{m}{m-1}\right)}} \right]^{\frac{1}{r}} \\
r = 0: \quad & s_0 \leq C \ln(\ln(T))^{-\frac{d+m-1}{dm+m-1}} \left[\frac{\sqrt{T}}{N^{\left(\frac{2}{d}+\frac{2}{m-1}\right)}} \right]^{\frac{1}{\left(\frac{1}{d}+\frac{m}{m-1}\right)}}, \\
& \lambda \geq C \ln(\ln(T))^{1/m} \frac{N^{1/m}}{\sqrt{T}}
\end{aligned}
\tag{B.1}
$$

and when N, T are sufficiently large, the following holds with probability at least $1 - \upsilon_{N,T}^{\mathcal{CC}} - \upsilon_{N,T}^{\mathcal{E}}$:

$$(i) \quad \frac{1}{T} \left\|\boldsymbol{X}(\hat{\boldsymbol{\beta}}^{(L)} - \boldsymbol{\beta})\right\|_2^2 \leq C\lambda^{2-r} s_r, \qquad (ii) \quad \left\|\hat{\boldsymbol{\beta}}^{(L)} - \boldsymbol{\beta}\right\|_1 \leq C\lambda^{1-r} s_r,$$

where $\upsilon_{N,T}^{\mathcal{CC}}$ and $\upsilon_{N,T}^{\mathcal{E}}$ are sequences that converge to 0 for large T, N (defined formally in the proof).

Proof: Let $P_\lambda := \{j : |\beta_j| > \lambda\}$, and $P_{\mathcal{S},\lambda} := \mathcal{S} \bigcup P_\lambda$, such that by construction, $\mathcal{S} \subseteq P_{\mathcal{S},\lambda}$. It then follows that by Assumption A.3 and Lemma B.2, we have on the set $\mathcal{E}_T(T\frac{\lambda}{4}) \bigcap \mathcal{CC}_T(P_{\mathcal{S},\lambda})$

$$\left\|\boldsymbol{X}(\hat{\boldsymbol{\beta}}^{(L)} - \boldsymbol{\beta})\right\|_2^2/T + \frac{\lambda}{4}\|\hat{\boldsymbol{\beta}}^{(L)} - \boldsymbol{\beta}\|_1 \leq C_1\lambda^2|P_{\mathcal{S},\lambda}| + C_2\lambda\|\boldsymbol{\beta}_{P_{\mathcal{S},\lambda}^c}\|_1.$$

It follows directly from Assumption A.2 that

$$|P_{\mathcal{S},\lambda}| \leq H + \sum_{j=1}^N \mathbb{1}_{\{|\beta_j|>\lambda\}} \left(\frac{|\beta_j|}{\lambda}\right)^r \leq H + \lambda^{-r}\sum_{j=1}^N |\beta_l|^r = H + \lambda^{-r}s_r.$$

Note that since $H \leq C$ by Assumption A.4 and both λ^{-r} and s_r asymptotically grow as $N, T \to \infty$, for sufficiently large N and T, $H + \lambda^{-r}s_r \leq C\lambda^{-r}s_r$. Following the proof of Lemma A.7 in Adamek et al. (2022b), we can also show that $\left\|\boldsymbol{\beta}_{P_{\mathcal{S},\lambda}^c}\right\|_1 \leq \lambda^{1-r}s_r$, and obtain the error bound

$$\frac{\|\boldsymbol{X}(\hat{\boldsymbol{\beta}}^{(L)} - \boldsymbol{\beta})\|_2^2}{T} + \lambda\|\hat{\boldsymbol{\beta}}^{(L)} - \boldsymbol{\beta}\|_1 \leq C_1\lambda^{2-r}s_r + C_2\lambda^{2-r}s_r = C\lambda^{2-r}s_r.$$

24

For the set $\mathcal{CC}_T(P_{\mathcal{S},\lambda})$, under Assumptions A.1 to A.3, we can apply Lemma A.3 in Adamek et al. (2022b) with $\eta_T = 1/\ln(\ln(T))^{\frac{dm+m-1}{d+m-1}}$ to show that $\mathbb{P}\left(\mathcal{CC}_T(P_\lambda)\right) \geq 1 - 3\left[1/\ln(\ln(T))\right]^{\frac{dm+m-1}{d+m-1}} := 1 - v_{N,T}^{\mathcal{CC}} \to 1$ as $N, T \to \infty$. Note that the condition $\eta_T \leq \frac{N^2}{e}$ is satisfied for sufficiently large N, T, and we also use this in the proof of Lemma A.3 in Adamek et al. (2022b) to plug in $|P_{\mathcal{S},\lambda}| \leq H + \lambda^{-r}s_r \leq C\lambda^{-r}s_r$.

Regarding the set $\mathcal{E}_T(T\frac{\lambda}{4})$, by Assumption A.1 and Lemma A.4 in Adamek et al. (2022b), $\mathbb{P}\left(\mathcal{E}_T(T\lambda/4)\right) \geq 1 - CN(\sqrt{T}\lambda)^{-m} := 1 - v_{N,T}^{\mathcal{E}}$ By the union bound,

$$\mathbb{P}\left(\mathcal{E}_T(T\lambda/4) \cap \mathcal{CC}_T(P_{\mathcal{S},\lambda})\right) \geq 1 - v_{N,T}^{\mathcal{E}} - v_{N,T}^{\mathcal{CC}},$$

and therefore the error bound holds with this probability as well. With the error bound, items (i) and (ii) follow straightforwardly. $\square$

Proof of Theorem 1: This result follows from Corollary 2 of Adamek et al. (2022b), with some small differences summarized here. First, in Adamek et al. (2022b), the tested hypotheses involve the full vector of parameters $\boldsymbol{\beta}$, whereas our result only holds for $\boldsymbol{\beta}_{\mathcal{S}}$. However, the matrix $\boldsymbol{R}_N$ in Adamek et al. (2022b) is restricted to have only H nonzero columns such that $\left\{j : \sum_{p=1}^{P} |r_{p,j}| > 0\right\} = \mathcal{S}$. The setup here is therefore equivalent to taking $\boldsymbol{R}_N = \left(\underset{S \times S}{\boldsymbol{I}}, \underset{S \times (N-S)}{\boldsymbol{0}}\right)$, and the elements we omit here are equal to zero.

Second, parts of the proof which require bounds on $\left\|\hat{\boldsymbol{\beta}}^{(L)} - \boldsymbol{\beta}\right\|_1$ or $\left\|\hat{\boldsymbol{u}}^{(L)} - \boldsymbol{u}\right\|_2$ need to be addressed by the new lasso error bound in Lemma B.3, rather than by Corollary 1 in Adamek et al. (2022b). Specifically, in Lemma B.8 of Adamek et al. (2022b) we need the bound $\left\|\hat{\boldsymbol{\beta}}^{(L)} - \boldsymbol{\beta}\right\|_1 \leq C\lambda^{1-r}s_r$, and in Lemma B.13 of Adamek et al. (2022b) we need $\|\hat{\boldsymbol{u}} - \boldsymbol{u}\|_2 = \left\|\boldsymbol{X}\left(\hat{\boldsymbol{\beta}}^{(L)} - \boldsymbol{\beta}\right)\right\|_2 \leq C\sqrt{T\lambda^{2-r}s_r}$. Note that since the new lasso error bound is the same as for the regular lasso, no changes to these proofs are necessary. $\square$

Appendix C   Simulations: Extra Figures
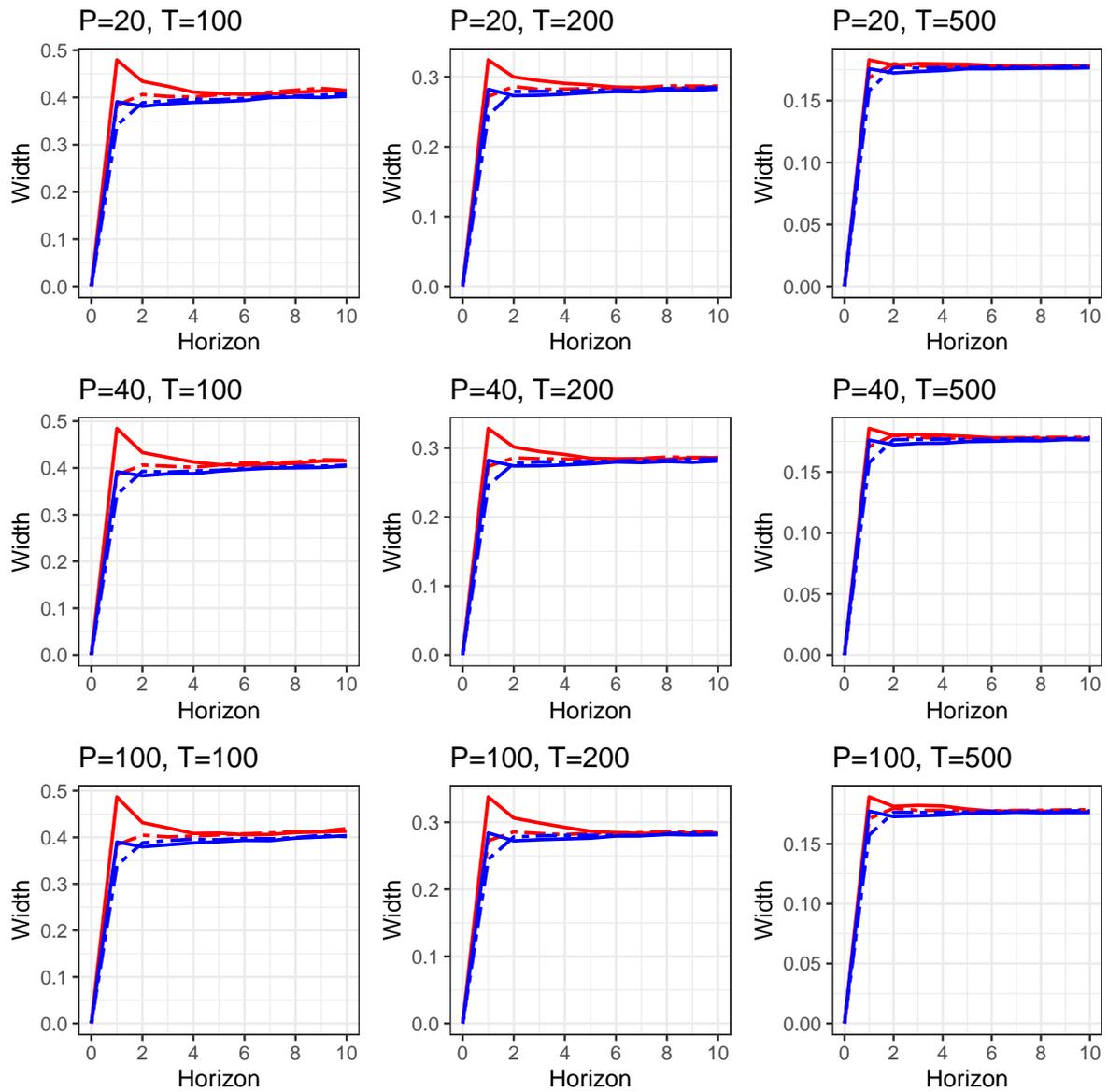
Figure C.1: Interval widths of the standard desparsified lasso (red) and the proposed desparsified lasso with ϕ_h unpenalized (blue). Dashed lines indicate results for the sign-switching DGP.

Appendix D Data Description

Table D.1: Definition of transformation codes

T	Transformation
1	$f(x_t) = x_t$
2	$f(x_t) = x_t - x_{t-1}$
3	$f(x_t) = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})$
4	$f(x_t) = \log(x_t)$
5	$f(x_t) = \log(x_t) - \log(x_{t-1})$
6	$f(x_t) = (\log(x_t) - \log(x_{t-1})) - (\log(x_{t-1}) - \log(x_{t-2}))$

In the following tables, the data transformation codes follow the definitions in Table D.1. Codes marked with an asterisk indicate a transformation that is different from the default provided by McCracken and Ng (2016) where we use the transformations in Bernanke et al. (2005) instead. The F/S column indicates whether we treat the variable as fast or slow respectively for our identification scheme.

Table D.2: Output & Income

	FRED	Description	DRI/McGraw	T	F/S
1	RPI	Real Personal Income	GMPYQ	5	S
2	W875RX1	Real personal income ex transfer receipts	GMYXPQ	5	S
3	INDPRO	IP Index	IP	5	S
4	IPFPNSS	IP: Final Products and Nonindustrial Supplies	IPP	5	S
5	IPFINAL	IP: Final Products (Market Group)	IPF	5	S
6	IPCONGD	IP: Consumer Goods	IPC	5	S
7	IPDCONGD	IP: Durable Consumer Goods	IPCD	5	S
8	IPNCONGD	IP: Nondurable Consumer Goods	IPCN	5	S
9	IPBUSEQ	IP: Business Equipment	IPE	5	S
10	IPMAT	IP: Materials	IPM	5	S
11	IPDMAT	IP: Durable Materials	IPMD	5	S
12	IPNMAT	IP: Nondurable Materials	IPMND	5	S
13	IPMANSICS	IP: Manufacturing (SIC)	IPMFG	5	S
14	IPB51222S	IP: Residential Utilities	IPUT	5	S
15	IPFUELS	IP: Fuels	-	5	S
17	CUMFNS	Capacity Utilization: Manufacturing	IPXMCA	1*	S

Table D.3: Money & Credit

	FRED	Description	DRI/McGraw	T	F/S
1	M1SL	M1 Money Stock	FM1	5*	F
2	M2SL	M2 Money Stock	FM2	5*	F
3	M2REAL	Real M2 Money Stock	FM2DQ	5	F
4	BOGMBASE	St. Louis Adjusted Monetary Base	FMFBA	5*	F
5	TOTRESNS	Total Reserves of Depository Institutions	FMRRA	5*	F
6	NONBORRES	Reserves Of Depository Institutions	FMRNBA	5*	F
7	BUSLOANS	Commercial and Industrial Loans	FCLNQ	5*	F
8	REALLN	Real Estate Loans at All Commercial Banks	-	6	F
9	NONREVSL	Total Nonrevolving Credit	CCINRV	5*	F
10	CONSPI	Nonrevolving consumer credit to Personal Income	-	2	F
12	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	-	6	F
13	DTCTHFNM	Total Consumer Loans and Leases Outstanding	-	6	F
14	INVEST	Securities in Bank Credit at All Commercial Banks	-	6	F

Table D.4: Labour Market

	FRED	Description	DRI/McGraw	T	F/S
1	HWI	Help-Wanted Index for United States	LHEL	5*	S
2	HWIURATIO	Ratio of Help Wanted/No. Unemployed	LHELX	4*	S
3	CLF16OV	Civilian Labor Force	LHEM	5	S
4	CE16OV	Civilian Employment	LHNAG	5	S
5	UNRATE	Civilian Unemployment Rate	LHUR	1*	S
6	UEMPMEAN	Average Duration of Unemployment (Weeks)	LHU680	1*	S
7	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	LHU5	1*	S
8	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	LHU14	1*	S
9	UEMP15OV	Civilians Unemployed - 15 Weeks & Over	LHU15	1*	S
10	UEMP15T26	Civilians Unemployed for 15-26 Weeks	LHU26	1*	S
11	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	-	5	S
12	CLAIMSx	Initial Claims	-	5	S
13	PAYEMS	All Employees: Total nonfarm	LPNAG	5	S
14	USGOOD	All Employees: Goods-Producing Industries	LPGD	5	S
15	CES1021000001	All Employees: Mining and Logging: Mining	LPMI	5	S
16	USCONS	All Employees: Construction	LPCC	5	S
17	MANEMP	All Employees: Manufacturing	LPEM	5	S
18	DMANEMP	All Employees: Durable goods	LPED	5	S
19	NDMANEMP	All Employees: Nondurable goods	LPEN	5	S
20	SRVPRD	All Employees: Service-Providing Industries	LPSP	5	S
21	USTPU	All Employees: Trade, Transportation & Utilities	LPTU	5	S
22	USWTRADE	All Employees: Wholesale Trade	LPT	5	S
23	USTRADE	All Employees: Retail Trade	-	5	S
24	USFIRE	All Employees: Financial Activities	LPFR	5	S
25	USGOVT	All Employees: Government	LPGOV	5	S
26	CES0600000007	Avg Weekly Hours : Goods-Producing	-	1	S
27	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	LPMOSA	1*	S
28	AWHMAN	Avg Weekly Hours : Manufacturing	LPHRM	1	S
30	CES0600000008	Avg Hourly Earnings : Goods-Producing	-	6	S
31	CES2000000008	Avg Hourly Earnings : Construction	LEHCC	5*	S
32	CES3000000008	Avg Hourly Earnings : Manufacturing	LEHM	5*	S

Table D.5: Consumption & Orders

	FRED	Description	DRI/McGraw	T	F/S
1	HOUST	Housing Starts: Total New Privately Owned	HSFR	4	F
2	HOUSTNE	Housing Starts, Northeast	HSNE	4	F
3	HOUSTMW	Housing Starts, Midwest	HSMW	4	F
4	HOUSTS	Housing Starts, South	HSSOU	4	F
5	HOUSTW	Housing Starts, West	HSWST	4	F
6	PERMIT	New Private Housing Permits (SAAR)	-	4	F
7	PERMITNE	New Private Housing Permits, Northeast (SAAR)	-	4	F
8	PERMITMW	New Private Housing Permits, Midwest (SAAR)	-	4	F
9	PERMITS	New Private Housing Permits, South (SAAR)	-	4	F
10	PERMITW	New Private Housing Permits, West (SAAR)	-	4	F

Table D.6: Orders & Inventories

	FRED	Description	DRI/McGraw	T	F/S
1	DPCERA3M086SBEA	Real personal consumption expenditures	GMCQ	5	S
2	CMRMTSPLx	Real Manu. and Trade Industries Sales	-	5	F
3	RETAILx	Retail and Food Services Sales	-	5	F
9	AMDMNOx	New Orders for Durable Goods	-	5	F
11	AMDMUOx	Unfilled Orders for Durable Goods	-	5	F
12	BUSINVx	Total Business Inventories	-	5	F
13	ISRATIOx	Total Business: Inventories to Sales Ratio	-	2	F

Table D.7: Interest rate & Exchange rates

	FRED	Description	DRI/McGraw	T	F/S
1	FEDFUNDS	Effective Federal Funds Rate	FYFF	1*	F
2	CP3Mx	3-Month AA Financial Commercial Paper Rate	-	2	F
3	TB3MS	3-Month Treasury Bill	FYGM3	1*	F
4	TB6MS	6-Month Treasury Bill	FYGM6	1*	F
5	GS1	1-Year Treasury Rate	FYGT1	1*	F
6	GS5	5-Year Treasury Rate	FYGT5	1*	F
7	GS10	10-Year Treasury Rate	FYGT10	1*	F
8	AAA	Moody's Seasoned Aaa Corporate Bond Yield	FYAAAC	1*	F
9	BAA	Moody's Seasoned Baa Corporate Bond Yield	FYBAAC	1*	F
10	COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS	-	1	F
11	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	SFYGM3	1	F
12	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS	SFYGM6	1	F
13	T1YFFM	1-Year Treasury C Minus FEDFUNDS	SFYGT1	1	F
14	T5YFFM	5-Year Treasury C Minus FEDFUNDS	SFYGT5	1	F
15	T10YFFM	10-Year Treasury C Minus FEDFUNDS	SFYGT10	1	F
16	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS	SFYAAAC	1	F
17	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS	SFYBAAC	1	F
19	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	EXRSW	5	F
20	EXJPUSx	Japan / U.S. Foreign Exchange Rate	EXRJAN	5	F
21	EXUSUKx	U.S. / U.K. Foreign Exchange Rate	EXRUK	5	F
22	EXCAUSx	Canada / U.S. Foreign Exchange Rate	EXRCAN	5	F

Table D.8: Prices

	FRED	Description	DRI/McGraw	T	F/S
1	WPSFD49207	PPI: Finished Goods	PWFSA	5*	S
2	WPSFD49502	PPI: Finished Consumer Goods	PWFCSA	5*	S
3	WPSID61	PPI: Intermediate Materials	PWIMSA	5*	S
4	WPSID62	PPI: Crude Materials	PWCMSA	5*	S
5	OILPRICEx	Crude Oil, spliced WTI and Cushing	-	6	F
6	PPICMM	PPI: Metals and metal products	-	6	S
8	CPIAUCSL	CPI : All Items	PUNEW	5*	S
9	CPIAPPSL	CPI : Apparel	PU83	5*	S
10	CPITRNSL	CPI : Transportation	PU84	5*	S
11	CPIMEDSL	CPI : Medical Care	PU85	5*	S
12	CUSR0000SAC	CPI : Commodities	PUC	5*	S
13	CUSR0000SAD	CPI : Durables	PUCD	5*	S
14	CUSR0000SAS	CPI : Services	PUS	5*	S
15	CPIULFSL	CPI : All Items Less Food	PUXF	5*	S
16	CUSR0000SA0L2	CPI : All items less shelter	PUXHS	5*	S
17	CUSR0000SA0L5	CPI : All items less medical care	PUXM	5*	S
18	PCEPI	Personal Cons. Expend.: Chain Index	-	6	S
19	DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	GMCDQ	5*	S
20	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	GMCNQ	5*	S
21	DSERRG3M086SBEA	Personal Cons. Exp: Services	GMCSQ	5*	S

Table D.9: Stock Market

	FRED	Description	DRI/McGraw	T	F/S
1	S&P 500	S&P's Common Stock Price Index: Composite	FSPCOM	5	F
2	S&P: indust	S&P's Common Stock Price Index: Industrials	FSPIN	5	F
3	S&P div yield	S&P's Composite Common Stock: Dividend Yield	FSDXP	1*	F
4	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio	FSPXE	1*	F

Appendix E   FAVAR Implementation

We closely follow the method described in Bernanke et al. (2005) to estimate the FAVAR. The most important difference concerns the scaling of all impulse responses by the response of the FFR at horizon 0. That means the shock to which the variables are responding is of such a size that the FFR rises by one on impact, as opposed to a size of one standard deviation, as is done by Bernanke et al. (2005). We implement this change to ensures that the scale of the responses from the FAVAR are comparable to those from our HDLP specification.

Let $\underset{720\times122}{\boldsymbol{X}_{\text{all}}}$ denote the matrix containing all variables, and $\underset{720\times67}{\boldsymbol{X}_{\text{slow}}}$ the one containing only the "slow" variables which we assume to not react within the same period to a monetary policy shock. We then estimate the factors $\underset{720\times3}{\hat{\boldsymbol{C}}}$ and $\underset{720\times3}{\hat{\boldsymbol{C}}^*}$ as the first 3 principal components of $\boldsymbol{X}_{\text{all}}$ and $\boldsymbol{X}_{\text{slow}}$ respectively. We regress $\hat{\boldsymbol{C}}$ onto $\hat{\boldsymbol{C}}^*$ and $\underset{720\times1}{\boldsymbol{R}_s}$, where $\boldsymbol{R}_s$ is the FFR ($\boldsymbol{R}$) scaled such that it has mean 0 and variance 1, and estimate by OLS

$$\hat{\boldsymbol{C}} = \left[\hat{\boldsymbol{C}}^*, \boldsymbol{R}_s\right] \underset{4\times3}{\hat{\boldsymbol{\beta}}} + \hat{\boldsymbol{u}}.$$

We then take $\underset{1\times3}{\hat{\boldsymbol{b}}_R}$ as the last row of $\hat{\boldsymbol{\beta}}$, and let $\underset{720\times3}{\hat{\boldsymbol{F}}} = \hat{\boldsymbol{C}} - \boldsymbol{R}_s\boldsymbol{b}_R$. Next, we use the `vars` package in R to estimate a 13 lag VAR with the four variables in $[\hat{\boldsymbol{F}}, \boldsymbol{R}]$ (ordering the FFR last), and obtain the impulse responses of the factors and $\boldsymbol{R}$ using a recursive identification scheme with Cholesky decomposition, and store them in a matrix $\underset{(h_{\max}+1)\times4}{\boldsymbol{IR}_{\text{fac}}}$. These impulse responses are all scaled such that we use the unit shock identification, similar to that of local projections. To obtain the impulse responses of the individual variables in $\boldsymbol{X}_{\text{all}}$, we regress $\boldsymbol{X}_{\text{all}}$ onto $\hat{\boldsymbol{F}}$, and estimate by OLS

$$\boldsymbol{X}_{\text{all}} = \hat{\boldsymbol{F}} \underset{3\times122}{\hat{\boldsymbol{\Lambda}}} + \hat{\boldsymbol{v}}.$$

The matrix of impulse responses for all variables is then obtained by

$$\underset{(h_{\max}+1)\times122}{\boldsymbol{IR}} = \boldsymbol{IR}_{\text{fac}} \begin{bmatrix} \hat{\boldsymbol{\Lambda}} \\ \\ \mathbf{0} \end{bmatrix}.$$

Finally, we cumulate the impulse responses for the variables which were taken in differences to obtain the response of the level variable.

For inference, we use a simple residual bootstrap. We draw with replacement from the VAR residuals, and use them to construct the bootstrap series $\hat{\boldsymbol{F}}^*$ and $\boldsymbol{R}^*$. We then perform the same steps, i.e. estimate the VAR on the bootstrap data, obtain (unit-shock scaled) impulse responses for the factors, compute the (cumulated) implied impulse responses of all variables. We repeat this process $B = 499$ times and use the 2.5% and 97.5% quantiles to construct confidence intervals.