Supplemental Appendix to "Linear Regression with Weak Exogeneity"

Anna Mikusheva*, Mikkel Sølvsten†

February 2025

This supplemental appendix contains additional theoretical results referenced in the main paper. Specifically, Section A contains a theoretical example in which we can derive the asymptotic bias as a function of primitives. Section 2 of the paper discusses this example immediately after Theorem 1. Section B explains how our framework extends to VAR models. Section 5 of the paper references this explanation in its discussion of extending the original framework to a setting with infinite lag feedback that decays relatively fast. Finally, Section C contains additional simulation studies that investigate the behavior of the proposed IV estimator in feedback models that deviate in various ways from the main setup considered in the paper and compare with additional estimators a researcher might have considered.

Appendix A Theoretical example

As a proof of concept, this section introduces a theoretical example where the average sample autocorrelation $\hat{\rho}$ of the matrix $\tilde{X}$ remains bounded away from zero in large samples. We construct a time series $\tilde{x}_t$ with short dependence in which the average population first-order auto-correlation equals $1/2$ and $\frac{1}{K}\sum_t \tilde{M}_{tt-1} \xrightarrow{p} \frac{-1}{2}$. For this example, we therefore show that the asymptotic bias in the worst direction is $\tau\frac{K}{T}$, where $\tau \neq 0$ is a constant depending on

*Department of Economics, M.I.T., 50 Memorial Drive, E52-526, Cambridge, MA, 02142, United States. E-mail: amikushe@mit.edu.

†Department of Economics and Business Economics, Aarhus University, Fuglsangs Allé 4, Building 2621, B 13a, 8210 Aarhus V, Denmark. E-mail: miso@econ.au.dk. This research was supported by grants from the Danish National and Aarhus University Research Foundations (DNRF Chair #DNRF154 and AUFF Grant #AUFF-E-2022-7-3)

the primitives of the data generating process. This example renders the OLS estimator for the parameter $\alpha'\beta$ inconsistent (when K is proportional to T) and standard t-statistic based inferences asymptotically invalid (when $K/\sqrt{T}$ is separated from zero asymptotically).

Example. Assume that $x_{1t} = \tilde{x}_{1t} + a\varepsilon_t$ and $x_{kt} = \tilde{x}_{kt}$ for $2 \leq k \leq K$, where $\tilde{x}_t = U_{\lfloor (t+1)/2 \rfloor}$ for $U_t \sim i.i.d. N(0, \sigma_U^2 I_K)$. The researcher observes $\{y_t, x_t\}_{t=1}^T$ with even T where $y_t = x_t'\beta + \varepsilon_t$. The errors $\{\varepsilon_t\}_{t=1}^T$ are *i.i.d.* with mean zero and variance σ^2 and are independent of $\tilde{X}$. We will derive that in this setting, as long as $K/T \to \phi$ with $\phi \in [0, 1/2)$, we have the following formula for the asymptotic bias of the OLS estimator for the first coefficient:

$$\hat{\beta}_1^{\text{OLS}} - \beta_1 = \tau \frac{K}{T} + o_p(1) \quad \text{where} \quad \tau = -\frac{\sigma^2}{2} \frac{a^2}{\sigma_U^2(1-2\phi) + a^2\sigma^2(1-\phi)}. \tag{1}$$

Specifically, if $\phi > 0$, then the OLS estimator for the first coefficient is inconsistent. If $K/\sqrt{T} \to \infty$, then the bias of the OLS estimator for the first coefficient asymptotically dominates its standard error, and the standard t-statistic based inferences are asymptotically invalid. The critical feature of this example is that we can derive a formula in terms of primitives for the leading term of the bias, which is non-random here.

The result follows from two statements we derive for this example:

$$\frac{1}{K} \sum_t \tilde{M}_{tt-1} \xrightarrow{p} \frac{-1}{2}, \tag{2}$$

$$\alpha'(\bar{S}/T)^{-1}\alpha \xrightarrow{p} \frac{a^2}{\sigma_U^2(1-2\phi) + a^2\sigma^2(1-\phi)}, \tag{3}$$

where the notation is as in Theorem 1.

The proof of (2) uses some ideas from Anatolyev and Smirnov (2024). First, note that

$$\sum_t \tilde{M}_{tt-1} = -\sum_{t=1}^T \tilde{P}_{tt+1} = -\sum_{j=1}^J U_j'(2U'U)^{-1}(U_j + U_{j+1}) = -\frac{K}{2} - \frac{1}{2}\sum_{j=1}^J P_{jj+1}^U,$$

where $P^U = U(U'U)^{-1}U'$ is a $J \times J$ projection matrix, $U = [U_1, \ldots, U_J]$, and $J = T/2$. Let

$\Sigma_j = \sum_{i \notin \{j,j+1\}}^{J} U_i U_i'$. The Sherman–Morrison formula gives us:

$$P^U_{jj+1} = \frac{U_j' \Sigma_j^{-1} U_{j+1}}{(1 + U_j' \Sigma_j^{-1} U_j)(1 + U_{j+1}' \Sigma_j^{-1} U_{j+1}) - (U_j' \Sigma_j^{-1} U_{j+1})^2}.$$

The denominator is always greater than or equal to one, so, $\mathbb{E}[P^U_{jj+1}]^2 \leq \mathbb{E}[U_j' \Sigma_j^{-1} U_{j+1}]^2$. The matrix Σ_j is independent from U_j and U_{j+1}, thus, $\mathbb{E}[U_j' \Sigma_j^{-1} U_{j+1}] = 0$ and $\mathbb{E}[U_j' \Sigma_j^{-1} U_{j+1}]^2 = \sigma_U^4 \mathbb{E} \operatorname{tr}(\Sigma_j^{-2})$. Hence,

$$\mathbb{E}\left[\frac{2}{K} \sum_j P^U_{jj+1}\right]^2 \leq \frac{4}{K^2} \sum_{j,i} \sqrt{\mathbb{E}[P^U_{jj+1}]^2} \sqrt{\mathbb{E}[P^U_{ii+1}]^2} \leq \frac{T^2}{K^2} \sigma_U^4 \mathbb{E} \operatorname{tr}(\Sigma_j^{-2}).$$

The matrix Σ_j has a Wishart distribution. Von Rosen (1988) derives a formula for the moments of the inverted Wishart distribution, and it implies in our setting that $\frac{T^2}{K^2} \mathbb{E} \operatorname{tr}(\Sigma_j^{-2}) \to 0$. Putting all derivations together, we arrive at the statement in (2).

To establish (3), we apply a formula for block inversion to matrix $\bar{S}$. We are interested only in the (1,1) element of this inverse:

$$[\bar{S}^{-1}]_{11} = \left(\tilde{X}_1' M_{-1} \tilde{X}_1 + a^2 \sigma^2 (T - K)\right)^{-1},$$

where the notation is the same as in Section 2.2. Using the structure of the regressors:

$$\frac{1}{T} \tilde{X}_1' M_{-1} \tilde{X}_1 = \frac{1}{J} U_1' M^U_{-1} U_1,$$

where $U = [U_1, U_{-1}]$ with U_1 a $J \times 1$ vector, and M^U_{-1} is the $J \times J$ matrix that projects off U_{-1}. By construction U_1 is independent from M^U_{-1}. Hence,

$$\mathbb{E}[\frac{1}{J} U_1' M^U_{-1} U_1] = \frac{\sigma_U^2}{J} \operatorname{tr}(M^U_{-1}) = \sigma_U^2 \frac{J - K + 1}{J} = \sigma_U^2 \frac{T - 2K + 2}{T}.$$

By standard arguments, we have

$$\mathbb{E}\left(\frac{1}{J}\sum_{i\neq j} M^U_{-1,ij}U_{1,i}U_{1,j}\right)^2 = \frac{2\sigma_U^4}{J^2}\sum_{i\neq j}(M^U_{-1,ij})^2 = O(1/J).$$

$$Var\left(\frac{1}{J}\sum_{i} M^U_{-1,ii}U_{1,i}^2\right) \leq \frac{3\sigma_U^4 \operatorname{tr}\left(M^U\right)}{J^2} = O(1/J).$$

These moment bounds finally imply that

$$\alpha'(\bar{S}/T)^{-1}\alpha = a^2[(\bar{S}/T)^{-1}]_{11} = \frac{a^2}{\sigma_U^2\frac{T-2K}{T} + a^2\sigma^2\frac{T-K}{T}}(1+o_p(1)),$$

and leads to statement (3) and hence (1).

Appendix B VAR models

This section discusses how a VAR model fits the baseline framework when extended to infinite feedback. Section 5 introduces a generalization of Assumption 1 by allowing for a finite number of feedback terms α_ℓ with $\ell = 1, ..., L$. Section 5 shows that the paper's main result, namely, the result on the OLS bias, generalizes, and the bias contains L terms corresponding to the appropriate feedback lags. We explicitly state that this result directly generalizes to the infinite feedback case as long as feedback size decays fast enough. This result, though a very natural generalization of the current paper, is technically demanding and deserving of a separate paper. Here, we only spell out how a typical VAR model widely used in empirical macroeconomics naturally fits this framework.

Consider a data-generating process described by a VAR(1) setup:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = A\begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + e_t. \tag{4}$$

Here, z_t is multi-dimensional and may include multiple lags of y_t and other variables while y_t is a scalar. It is customary to assume that e_t is an *i.i.d.* sequence of shocks (or forecast errors). One can recast any multi-dimensional VAR(p) as an (even higher dimensional) VAR(1), which is known as the companion-form representation. Stationarity of the VAR holds when $\|A\| < 1$. The literature typically estimates VAR models via equation-by-equation OLS. Consider OLS estimation of the first equation:

$$y_t = \beta' \underbrace{(y_{t-1}, z'_{t-1})'}_{=x_t} + e^y_t = x'_t \beta + e^y_t,$$

where β' is the first row of the matrix A, and e^y_t is the first element of e_t.

Assume that e_t is Gaussian and decompose it as $e_t = (1, \alpha')' e^y_t + (0, \xi'_t)'$, where ξ_t's dimension is one lower than e_t and ξ_t is independent from e^y_t. One implication of (4) is that

$$x_t = \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} = \underbrace{\sum_{j=0}^{\infty} A^j \begin{pmatrix} 0 \\ \xi_{t-j-1} \end{pmatrix}}_{=\tilde{x}_t} + \sum_{j=0}^{\infty} \underbrace{A^j \begin{pmatrix} 1 \\ \alpha \end{pmatrix}}_{\alpha_j} e^y_{t-j-1}$$

where $\tilde{x}_t$ is independent of e^y_t at all lags and leads, and thus, we can take it as the strictly exogenous part for the first equation. Here, we have a model with infinite feedback $\alpha_j = A^j(1, \alpha')'$, where the feedback size is geometrically decaying: $\|\alpha_j\| \leq \|A\|^j \sqrt{1 + \|\alpha\|^2}$. Thus, the stated requirement of a fast (here, geometrically) decaying feedback holds.

Appendix C Additional simulations

This section contains additional simulation results that support the paper's central message and explore the results' robustness in several directions.

C.1 Artificial data simulations

Short-run vs long-run dependence. Footnote 4 of the paper references this first simulation exercise. The simulation design underlying Figure 1 of the paper uses K independent AR(1) processes to generate $\tilde{X}$. Figure 1 shows that bias increases quickly with the autoregressive parameter ρ. In an AR(1) process, the parameter ρ characterizes both the short-term dependence and the long-run persistence, so one may wonder which of these features is essential for the result. Theorem 1 of the paper states that the first-order sample autocorrelation matters. We, therefore, repeat the experiment presented in Figure 1 of the paper but now simulate $\tilde{X}$ as independent MA(1) processes. In an MA(1) process, the parameter ρ characterizes the short-term dependence only.

The outcome vector is generated as $y = X\beta + \varepsilon$ with $\varepsilon \sim N(0, I)$ and $\beta = 0$. The design matrix is generated as $x_{1t} = \tilde{x}_{1t} + a\varepsilon_{t-1}$ and $X_{-1} = \tilde{X}_{-1}$, where $\tilde{X}$ is generated as a rotated MA(1) process with $\tilde{X}\tilde{X}'/T = I_K$, independent from ε. Specifically, we generate $v_t = \rho u_{t-1} + u_t$ with $\{u_t\}_{t=1}^T$ $i.i.d.$ $N(0, I_K)$ and define $\tilde{X} = V(V'V/T)^{-1/2}$, where the square root comes from Cholesky decomposition. Across simulations, we fix the sample size at $T = 200$ and the coefficient on the feedback mechanism at $a = 1.5$. Simulation results are summarized in Figure C.1 with the left panel showing results for the number of regressors K between 4 and 150 (fixing ρ at 0.8). The right panel reports the results for the autocorrelation in regressors ρ between 0 and 0.98 (fixing K at 50). We report simulated values of absolute bias and standard deviation for the first coordinate of OLS and IV together with the mean absolute value of the ratio of the lower trace of M to the sample size. The results present sixth-order polynomial fits to the simulation results across K. The results are extremely similar to those reported in Section 2 regarding the size of the bias/standard deviations and dependence on the number of regressors and their one-period predictability.

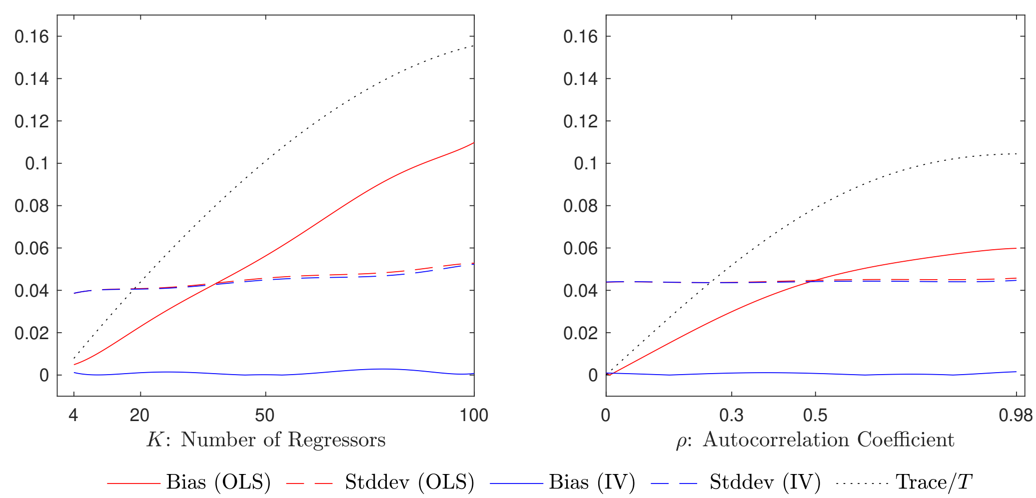

Figure C.1: Absolute Bias and Standard Deviation of OLS and IV with T=200

Inconsistency of OLS. We now report another experiment using artificial data, replicating the simulations reported in Figure 1 of the paper but for a larger sample size. Figure

C.2 presents results for the same simulation design formulated in Section 2.2 of the paper but with a sample size of $T = 800$. The number of regressors varies from 16 to 400. Here, the bias reaches the same level as in Figure 1 when the number of regressors is the same fraction of the sample size, while the standard deviations drop two-fold. These comparative statics demonstrate the inconsistency of the OLS for the worst direction when the number of regressors K grows proportionally to T. In essence, the estimator concentrates on an incorrect value as the sample size increases.

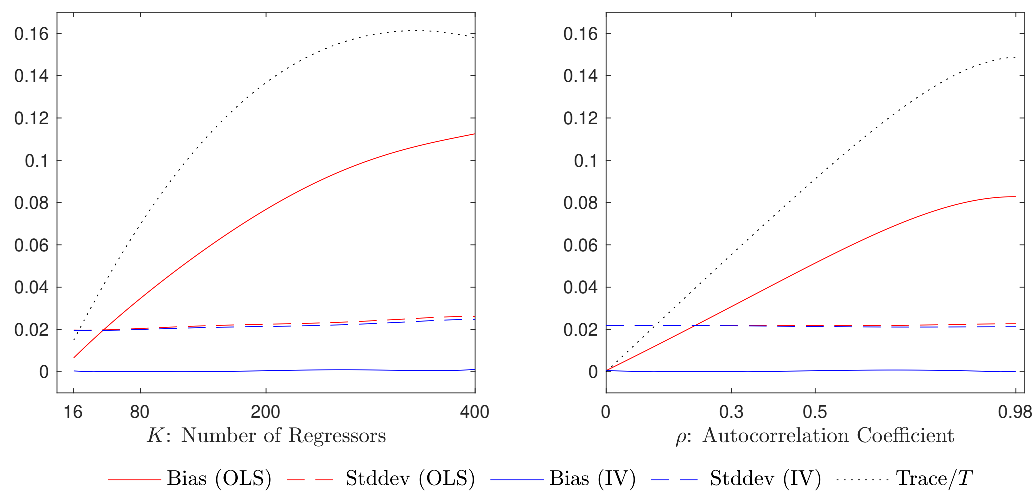

Figure C.2: Absolute Bias and Standard Deviation of OLS and IV with $T = 800$

C.2 Robustness to violations of Assumption 1

Our main results (OLS bias and the consistency of our newly proposed estimator) rely heavily on Assumption 1. This section explores how these results adjust when we relax parts of Assumption 1. The simulation experiment in this subsection tries to mimic the US macro data described in Section 6.

Two-period feedback We consider a violation of Assumption 1, part (i), by introducing two periods of empirically motivated feedback effects. Except for the additional feedback period, the simulation setup is identical to the baseline (homoskedastic) simulations in Section 6. Specifically, we simulate samples as $X = X_{\mathrm{r}} + D'\varepsilon\alpha_1' + (D')^2\varepsilon\alpha_2'$ and $y = X\beta + \varepsilon$ where $\alpha_1 = X_{\mathrm{r}}'D'e/(e'e)$ and $\alpha_2 = X_{\mathrm{r}}'(D')^2e/(e'e)$, $\varepsilon \sim N(0, \sigma^2 I)$, $\sigma^2 = e'e/(T-K)$ for $e = y_{\mathrm{r}} - X_{\mathrm{r}}\beta$.

We report results for estimators of the linear contrasts $\theta_1 = \alpha_1'\beta$ and $\theta_2 = (\alpha_1 + \alpha_2)'\beta$. The left panels of Figures C.3 and C.4 depict the results of the experiments (for different K) at the 10th percentile of the OLS bias. The right panels of Figures C.3 and C.4 contain the results of the experiments at the 90th percentile. For those experiments, Figure C.3 reports the OLS bias and standard deviation and the IV bias and standard deviation in the first feedback direction ($\theta_1 = \alpha_1'\beta$) along with the normalized lower trace of $M_{\mathrm{r}} = I - X_{\mathrm{r}}(X_{\mathrm{r}}'X_{\mathrm{r}})^{-1}X_{\mathrm{r}}'$ that is, $\mathrm{tr}(D'M_{\mathrm{r}})/T$. Figure C.4 reports similar indicators in the direction mixing the first and second feedback directions ($\theta_2 = (\alpha_1 + \alpha_2)'\beta$).

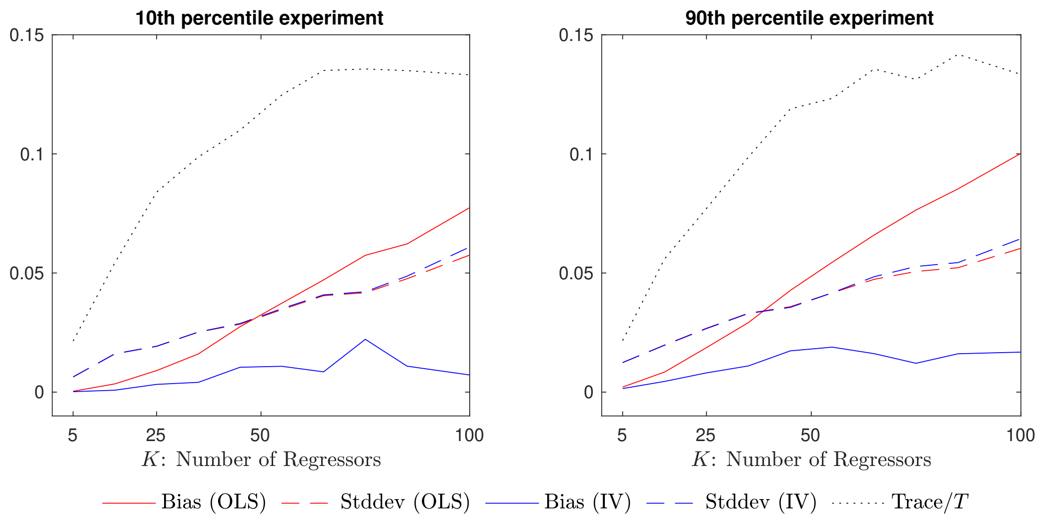

Figure C.3: Absolute Bias and Standard Deviation of OLS and IV with two-period feedback and $\theta_1 = \alpha_1'\beta$

Theorem 7 of the paper extends our results to multi-period feedback and shows that the bias of OLS has two terms corresponding to the two periods of feedback. It also shows that an IV-type estimator has no bias if its two lower traces are zero. The estimator we introduced relies heavily on the assumption of one-period feedback by only zeroing out the first lower diagonal. This simulation exercise aims to answer what happens when one misjudges the feedback's lag length. Figure C.3 shows that our IV estimator successfully corrects the bias in the direction of the first-order feedback, as predicted by Theorem 7. However, the bias due to the second lag feedback is still present, as shown in Figure C.4. There, the IV estimator has a noticeably smaller bias than the OLS, but some remain. Figures C.5 and C.6 report the size of t-statistic based OLS and IV inferences in the same experiments. The

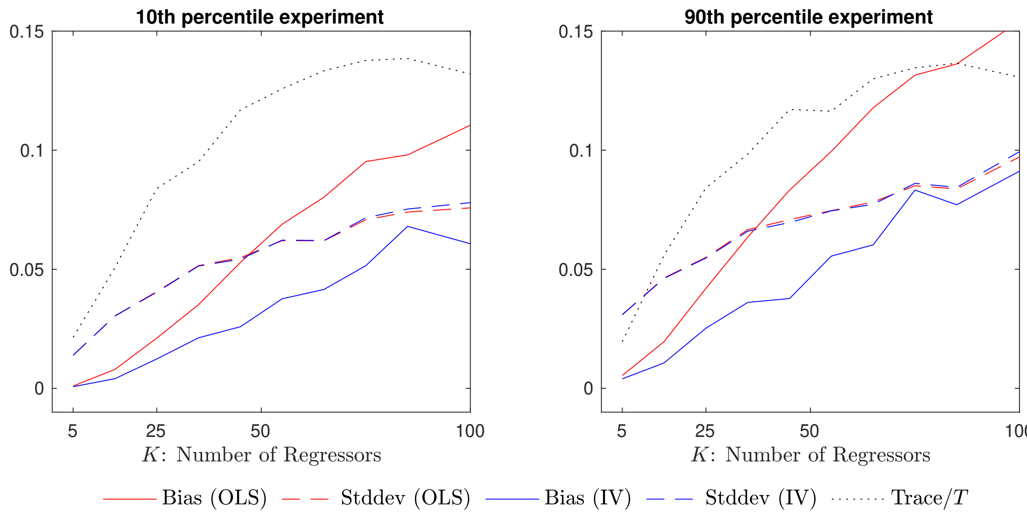

Figure C.4: Absolute Bias and Standard Deviation of OLS and IV with two-period feedback and $\theta_2 = (\alpha_1 + \alpha_2)'\beta$

results of Theorems 5 and 6 (on Gaussianity of estimators) extend to de-meaned versions of IV-type estimators that are not unbiased. One can see that in both cases of $\theta_1 = \alpha_1'\beta$ and $\theta_2 = (\alpha_1 + \alpha_2)'\beta$ the sizes depicted in Figures C.5 and C.6 trace extremely closely the biases of the corresponding estimators depicted in Figures C.3 and C.4, pointing out that the size distortions come from biases and not from violations of Gaussianity or inappropriate standard errors.

C.3 Robustness of inference

Non-Gaussian errors The asymptotically valid inference based on the t-statistic for the new estimator is established in Theorems 5 and 6 of the paper. Theorem 5 considers cases with a moderate number of regressors $K/T \to 0$ and uses only Assumption 1. Theorem 6 applies to the instances where the number of regressors is proportional to the sample size and assumes that the regression errors are Gaussian. The paper discusses that this Gaussianity of errors assumption matters for the standard errors. Skewed errors may require correcting the standard errors, though the correction is likely minor. We consider a violation of the assumption of Gaussian errors to explore whether inference is sensitive to this restriction. The simulation setup here is identical to the setup in Section 6 using the US data with one change: we generate mutually independent errors with $\varepsilon_t \mid V_t = v_t \sim sN(v_t, \sigma^2)$ and

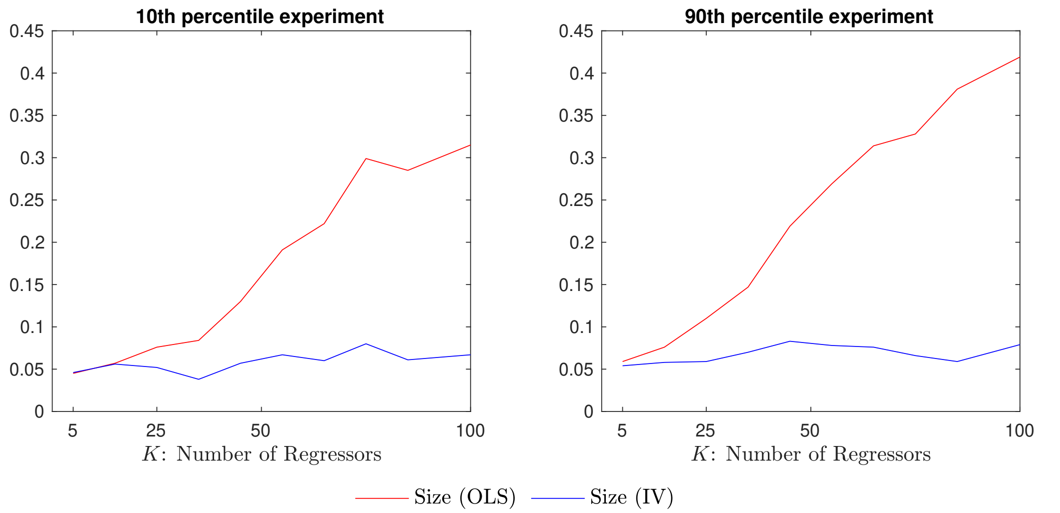

Figure C.5: Size of Nominal 5% two-sided tests using OLS and IV with two-period feedback and $\theta_1 = \alpha_1'\beta$

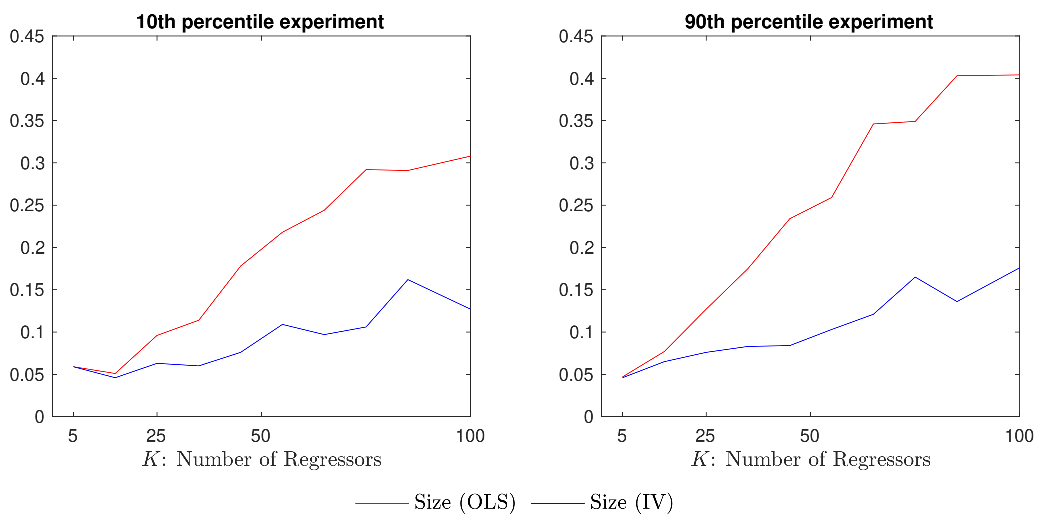

Figure C.6: Size of Nominal 5% two-sided tests using OLS and IV with two-period feedback and $\theta_2 = (\alpha_1 + \alpha_2)'\beta$

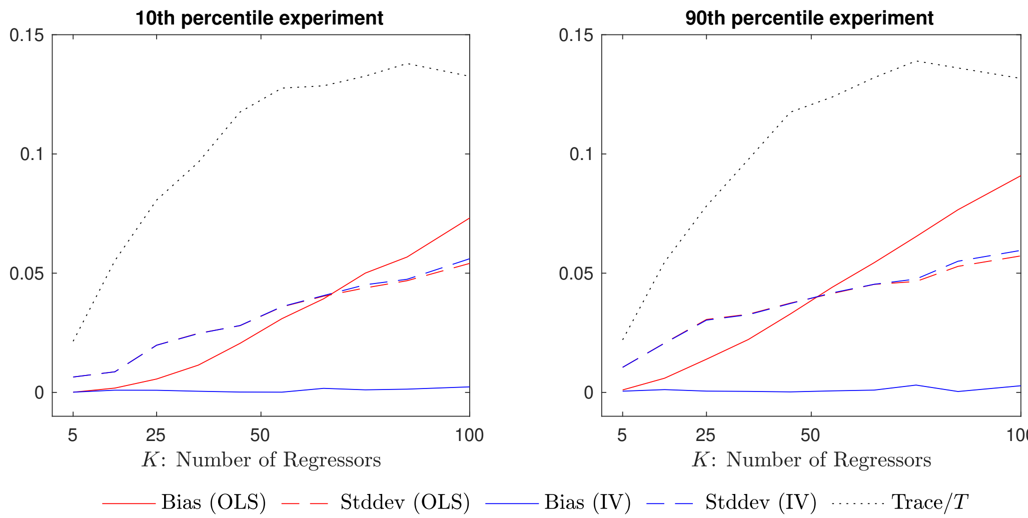

Figure C.7: Absolute Bias and Standard Deviation of OLS and IV with skewed errors

$V_t = -.6 + binomial(.6)$, where s is such that the variance of ε_t is σ^2. Thus, the errors follow an asymmetric 0.4/0.6 mixture of two Gaussian distributions (one with positive and one with negative mean).

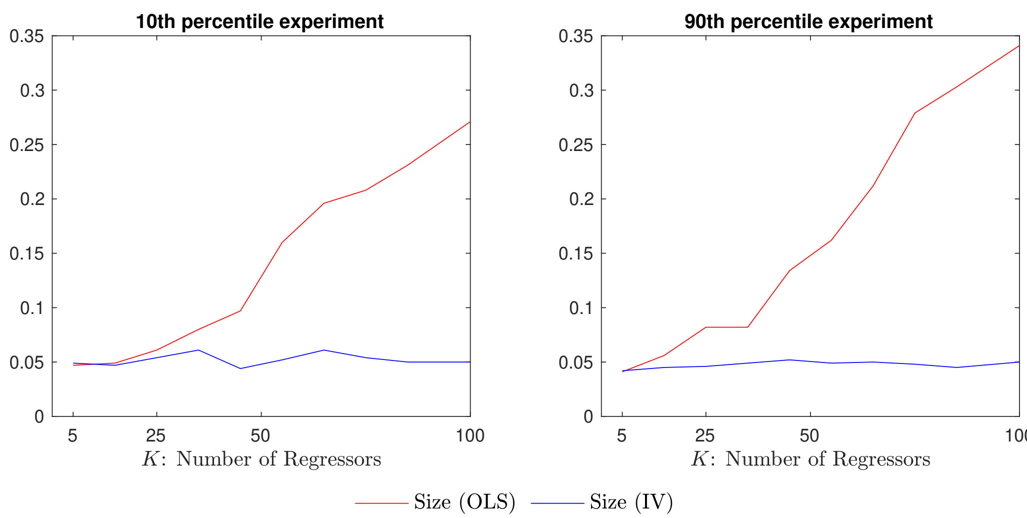

Figure C.8: Size of Nominal 5% two-sided tests using OLS and IV with skewed errors

Figure C.7 reports the biases and standard deviations of the OLS and IV estimators for the 10% and 90% experiments (the description is similar to the one in Figure 2 of the paper). The results are highly comparable to the ones reported there, as this non-Gaussian experimental design satisfies Assumption 1. Figure C.8 reports the size of tests based on the t-statistics for the same experiments. Our inference procedure controls size exceptionally well for all K, even without Gaussian errors.

Heteroskedasticity In this part, we consider a violation of Assumption 1, part (ii), by introducing empirically motivated heteroskedasticity. Here, the simulation design is precisely the heteroskedastic design described in Section 6 and underlying Figure 4 of the paper. Figure 4 of the paper depicts the biases and standard deviations of the OLS and the IV estimator when Assumption 1 is violated by conditional heteroskedasticity of unspecified form following the empirically observed one. We see from Figure 4 that bias correction is performed successfully in this case. Figure C.9 reports simulated size in the experiments (for different K) falling in the 10th and 90th percentile of the OLS size for the same experiments. Figure C.9 shows that despite an excellent bias correction, the inference based on the IV estimator t-statistic is imperfect. This issue arises because our standard errors are homoskedasticity-only and are not heteroskedasticity-robust.

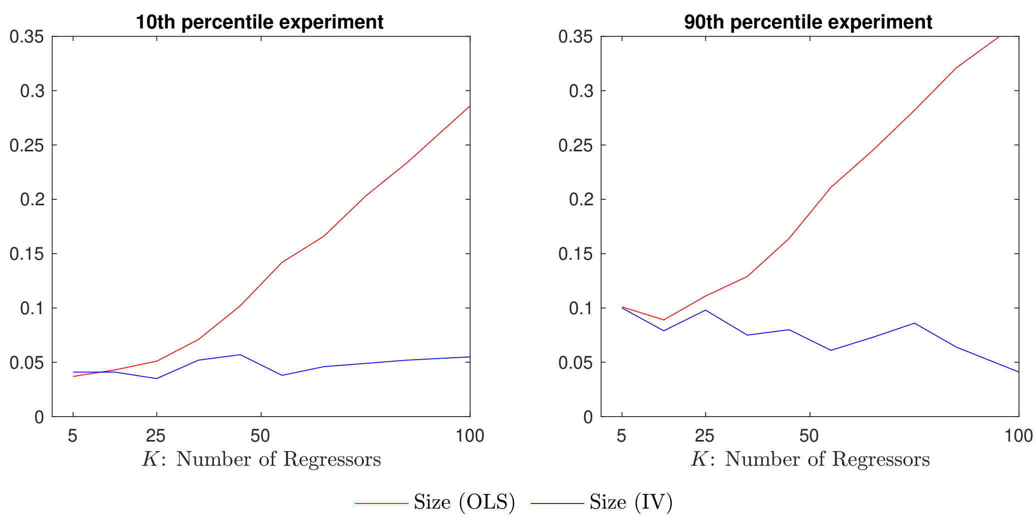

Figure C.9: Size of Nominal 5% two-sided tests using OLS and IV with heteroskedasticity

C.4 Alternatives to our estimator

Principal component analysis (PCA) One common suggestion in empirical macroeconomics when dealing with many regressors is to use PCA for dimension reduction. This suggestion relies on the empirical observation that a large set of macro indicators have a strong factor structure present. The essence of the suggestion is the following: one wants to estimate a regression model $y_t = x_t\beta + \varepsilon_t$ with interest in the coefficient β_1. To reduce the number of regressors, one first conducts a PCA on the $T \times (K-1)$ set of regressors X_{-1} to estimate a $T \times \hat{r}$ matrix of factors $[\hat{F}_1, ...\hat{F}_t]'$, where $\hat{r}$ is much smaller than T and always smaller than K. Then, one uses OLS to estimate the regression $y_t = x_{1t}\beta_1 + \hat{F}_t'\delta + v_t$, which has a relatively smaller number of regressors.

To assess the potential benefits of the PCA approach, we consider a simulation experiment that mimics the MA(1) design in Section C.1. The main difference from the baseline design is that the regressors have a strong factor structure, and the factors directly explain variation in the outcome. A key element that determines the potential omitted variable bias in PCA is whether a small number of factors capture most of the influence of $(x_{2t}, \dots, x_{Kt})$ on x_{1t} or y_t.

The structure of the regressors are

$$x_t = \Lambda F_t + \xi_t + \alpha\varepsilon_{t-1}$$

where Λ, $\{F_t\}_t$, $\{\xi_t\}_t$, and $\{\varepsilon_t\}_t$ are independent. The factors follow an MA(1) process: $F_t = \eta_t + \rho\eta_{t-1}$ and $\eta_t \sim i.i.d.N(0, \tau^2 I_r)$ where τ^2 is such that $\mathbb{E}[\|F_t\|^2] = r/(1+r)$. The idiosyncratic part of the regressors also follows an MA(1) process: $\xi_t = e_t + \rho e_{t-1}$ and $e_t \sim i.i.d.N(0, \tau^2 I_K)$. The factor loadings are orthogonal: $\Lambda = V(V'V/K)^{-1/2}$ where V is a $K \times r$ matrix with independent standard Gaussian entries. Finally, $\varepsilon \sim N(0, I)$ and $\alpha = (3a/4, a/4, 0, \dots, 0)'$, which spreads the feedback over two variables and induces a positive correlation between x_{1t} and x_{2t}. We use $a = 1.5$ and $\rho = 0.8$ as in Section C.1, while $r = 5$ ensures that the factors explain about 80% of the variation in the regressors.

The outcome equation is

$$y_t = x_{1t}\beta_1 + x_{2t}\beta_2 + F_t\delta + \varepsilon_t$$

where $\beta_1 = 0$, $\beta_2 = -0.5$, and $\delta = (1, \dots, 1)'/r$.

Figure C.10 shows the absolute bias and standard deviation of OLS, IV, and PCA as we vary the number of regressors, K. In our simulations, the PCA method uses the correct (so-called "oracle") number of factors, $\hat{r} = r = 5$, providing an advantage to PCA and

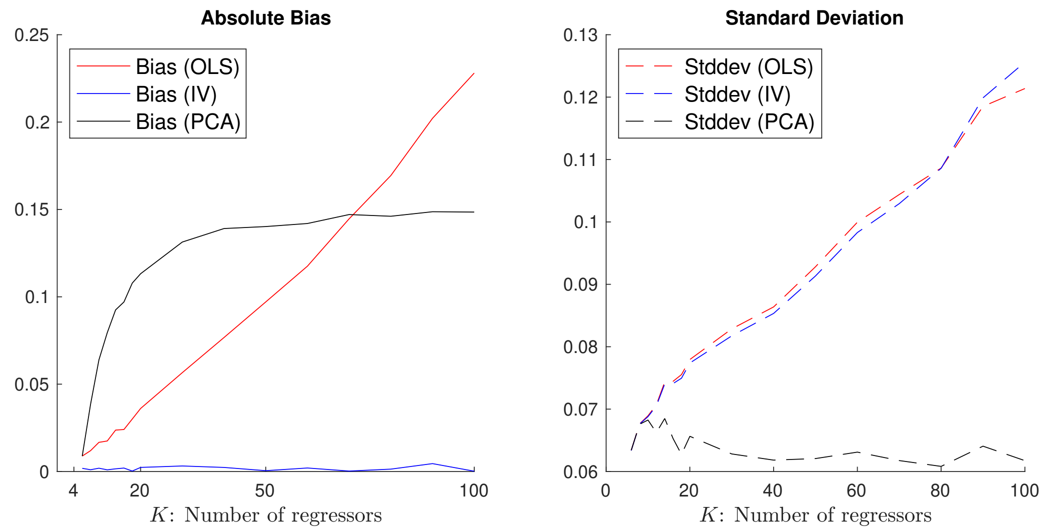

Figure C.10: Absolute Bias and Standard Deviation of OLS, IV, and PCA

removing additional uncertainty associated with the model selection step.[1] Figure C.10 demonstrates the inferior performance of PCA; it produces an estimator with a substantial bias and relatively small standard deviations. In other words, the estimator concentrates on the wrong value. The emerging bias of PCA is an omitted variable bias. Indeed, from an extensive literature on regressions with a rich set of controls, we know that a data compression approach (a.k.a. Machine Learning approach) would deliver valid results if it leads to good approximations in either the outcome equation or a regression of the regressor of interest (here x_{1t}) on the controls. In the current simulation setup, the factors capture neither the effect of controls on the outcome y_t nor the impact of controls on x_{1t}. We can see that knowing the identity of the regressor x_{2t} and placing it in the regression would successfully control the omitted variable bias. However, in an actual empirical application, we rarely know the exact identity of the needed controls, which is one of the main reasons for having many regressors. We set up the simulations to favor PCA since the impact of all regressors except the first two is captured well by the factor models. It is easy to see that by changing the coefficient β_2, we can change the size of the omitted variable bias.

Leave-one-out ideas As Section 2.2 explains, the OLS bias arises from the partialling out of many controls, which mixes in the lead of the weakly exogenous regressor and infects

[1] We also considered $\hat{r} = r + 1 = 6$ which lead to essentially the same results.

it with the current error term. An industrious reader might guess that a leave-one-out estimator that partials out the regressor in period t while dropping the observation from period $t+1$ potentially could restore consistency with many regressors. For estimation of β_1, this thinking leads to the estimator

$$\hat{\beta}_1^{\mathrm{LO}} = \frac{\sum_t (x_{1t} - \hat{x}_{1t,(t+1)}) y_t}{\sum_t (x_{1t} - \hat{x}_{1t(t+1)}) x_t}$$

where $\hat{x}_{1t,(t+1)}$ is the OLS prediction of x_{1t} using $(x_{2t}, \dots, x_{Kt})'$ in the sample of all observations but $t+1$. By the Sherman-Morrison-Woodbury formula, we have the following representation that highlights a relation between the IV estimator proposed in this paper and $\hat{\beta}_1^{\mathrm{LO}}$:

$$\hat{\beta}_1^{\mathrm{LO}} = \frac{\sum_t \left[(x_{1t} - \hat{x}_{1t}) - \frac{M^*_{t,t+1}}{M^*_{t+1,t+1}} (x_{1,t+1} - \hat{x}_{1,t+1}) \right] y_t}{\sum_t \left[(x_{1t} - \hat{x}_{1t}) - \frac{M^*_{t,t+1}}{M^*_{t+1,t+1}} (x_{1,t+1} - \hat{x}_{1,t+1}) \right] x_t}$$

where $\hat{x}_{1t}$ is the full sample OLS prediction. Specifically, when viewed as an IV estimator, we see that $\hat{\beta}_1^{\mathrm{LO}}$ use the partialled out regressor from the period ahead, $x_{1,t+1} - \hat{x}_{1,t+1}$, to offset some of the endogeneity in $x_{1t} - \hat{x}_{1t}$ which gets introduced by partialling out. However, this simple idea cannot ensure consistency at the same level of generality as the IV estimator. We do not delve into the deeper theoretical underpinnings of this failure but illustrate it by applying the estimator in the same simulations as in Section 6.

We report results for estimators of the linear contrasts $\theta = \alpha'\beta$. The left panel of Figure C.11 depicts the results of the experiments (for different K) at the 10th percentile of the bias for the leave-out estimator. The right panel of Figure C.11 contains the results of the experiments at the 90th percentile. We do see that while the leave-one-out idea may work in some circumstances, it does not always perform well.

Lag augmentation One referee proposed the following idea for the OLS bias corrections, and we are very grateful for the suggestion. The thought arises from an observation that if one had a good proxy for the previous period error term, then including it in the regression as a control would solve the weak exogeneity problem for the original regressors. In the absence of such a proxy, one may instead control for one lag of the outcome variable and

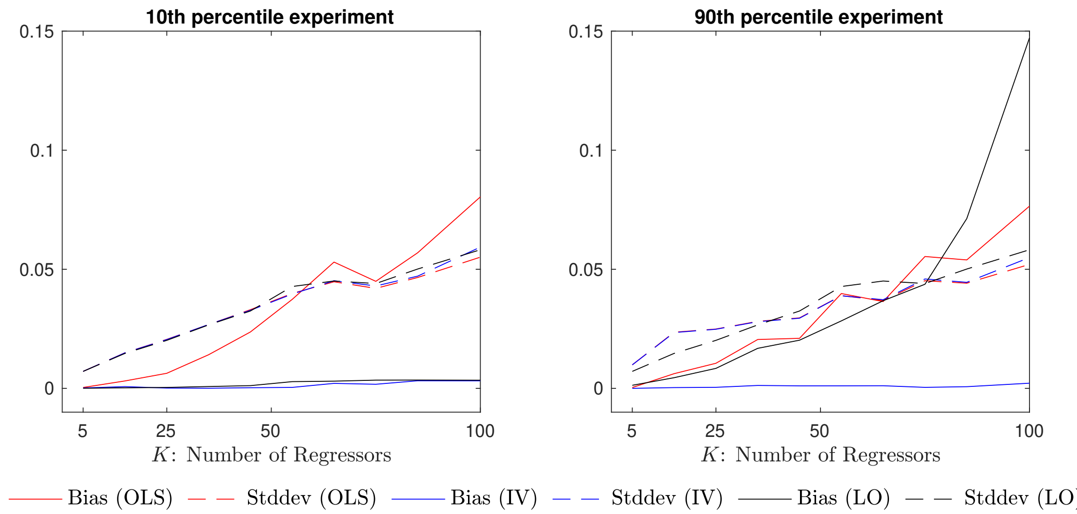

Figure C.11: Absolute Bias and Standard Deviation of OLS, IV, and Leave-Out

additionally one lag of all the regressors (that is, to run a regression of y_t on x_t, y_{t-1}, x_{t-1}). This adjustment more than doubles the number of regressors and ultimately leads to lower precision, but it would hopefully resolve the bias issue. Similarly, an alternative is to run an original OLS, calculate the residuals, and then re-run the OLS, including the lagged residual as an additional control.

We implemented both suggestions in the simulation setup underlying Figure 1 of the paper and described in Section 2.2. Specifically, the exogenous part of the regressors follows a K-dimensional AR(1) process with coefficient $\rho = 0.8$. The only change we made in comparison to Figure 1 is to use a non-zero value of β. The biases of OLS and our proposed estimator depicted in Figure 1 are invariant to the true value of β; however, for the proposal discussed in this section, the value of β seems to matter. We use $\beta = (0, \frac{1}{\sqrt{K}}, ..., \frac{1}{\sqrt{K}})'$. We calculate the following estimators: the OLS (red), our proposed IV (blue), the OLS of y_t on x_t, y_{t-1}, x_{t-1} (black), and the two-step OLS, where the first step calculates the residuals $\hat{e}_t$ from a regression of y_t on x_t, and then run OLS of y_t on x_t and $\hat{e}_{t-1}$ (magenta).

Figure C.12 reports the results of this exercise. We observe that controlling for the lags of both the outcome and the regressors does not work at all; it leads to increases in bias and variance of the estimator. Controlling for the lagged residual seems to correct the original OLS bias somewhat (though not entirely) at the expense of higher standard deviations. Our proposed estimator dominates the OLS with lagged residuals in terms of both bias and

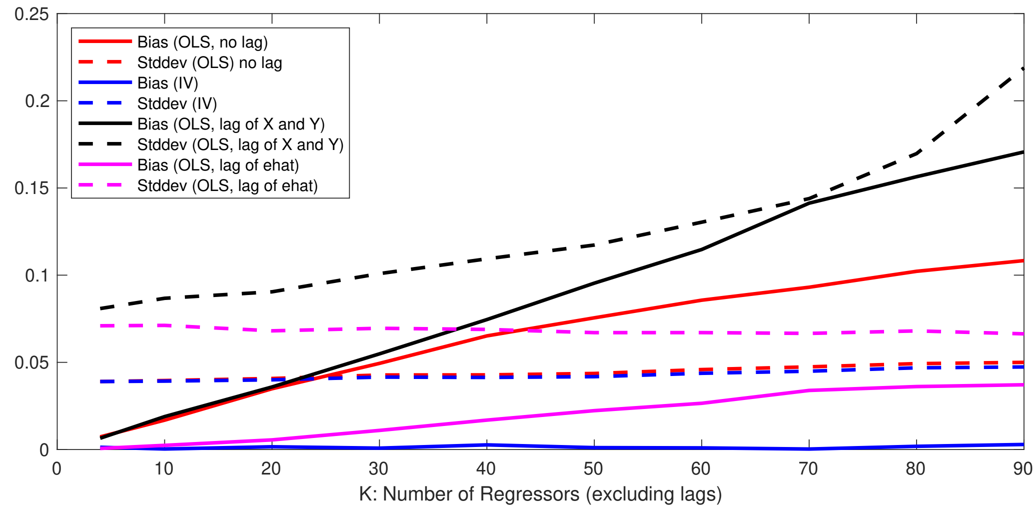

Figure C.12: Absolute Bias and Standard Deviation of OLS, IV, and OLS using lagged variables as additional controls

variance.

References

Anatolyev, S. and M. Smirnov (2024). Off-diagonal elements of projection matrices and dimension asymptotics. *Economics Letters 239*, 111761.

Von Rosen, D. (1988). Moments for the inverted Wishart distribution. *Scandinavian Journal of Statistics*, 97–109.