BIROn - Birkbeck Institutional Research Online

Pesaran, M.H. and Smith, Ron P. (2018) A Bayesian analysis of linear regression models with highly collinear regressors. Econometrics and Statistics 11 , pp. 1-21. ISSN 2452-3062.

Downloaded from: https://eprints.bbk.ac.uk/id/eprint/24824/

Usage Guidelines:
Please refer to usage guidelines at https://eprints.bbk.ac.uk/policies.html or alternatively contact lib-eprints@bbk.ac.uk.

A Bayesian Analysis of Linear Regression Models with Highly Collinear Regressors

M. Hashem Pesaran
USC Dornsife INET, University of Southern California,
USA and Trinity College, Cambridge, UK

Ron P. Smith*
Birkbeck, University of London

6 October 2018

Abstract

Exact collinearity between regressors makes their individual coefficients not identified. But, given an informative prior, their Bayesian posterior means are well defined. Just as exact collinearity causes non-identification of the parameters, high collinearity can be viewed as weak identification of the parameters, which is represented, in line with the weak instrument literature, by the correlation matrix being of full rank for a finite sample size T, but converging to a rank deficient matrix as T goes to infinity. The asymptotic behaviour of the posterior mean and precision of the parameters of a linear regression model are examined in the cases of exactly and highly collinear regressors. In both cases the posterior mean remains sensitive to the choice of prior means even if the sample size is sufficiently large, and that the precision rises at a slower rate than the sample size. In the highly collinear case, the posterior means converge to normally distributed random variables whose mean and variance depend on the prior means and prior precisions. The distribution degenerates to fixed points for either exact collinearity or strong identification. The analysis also suggests a diagnostic statistic for the highly collinear case. Monte Carlo simulations and an empirical example are used to illustrate the main findings.

JEL Classifications: C11, C18

Key Words: Bayesian identification, multicollinear regressions, weakly identified regression coefficients, highly collinear regressors.

*Corresponding author. Email address: r.smith@bbk.ac.uk.

1 Introduction

This paper presents a Bayesian analysis of the multicollinearity problem for stable linear regression models with highly collinear regressors. Multicollinearity is an old problem in time series analysis where the regressors tend to be highly persistent. For example, Spanos and McGuirk (2002, 365-6) note that although high degree of collinearity amongst the regressors is one of the recurring themes in empirical time series research, the manifestation of the problem seems unclear; there is no generally accepted way to detect it; and there is no generally accepted way to deal with it. Pesaran (2015, Section 3.11) discusses the multicollinearity problem and shows that in the case of highly collinear regressors the outcomes of individual t-tests and associated joint F-tests could be in conflict, with statistically insignificant outcomes for the individual t-test and a statistically significant outcome for the joint test. The term "multicollinearity" originates with Ragnar Frisch (1934) as a contraction of his phrase multiple collinearity which refers to a situation in which several linear relationships hold between variables and the meaning subsequently changed to linear dependence between regressors.

The adverse effects of multicollinearity on the precision with which the parameters are estimated can be reduced by the use of extra information, should it be available. The extra information can take the form of either more data or prior information. The prior information may be exact, for instance that a coefficient is zero or takes a particular value, or probabilistic, as in the Bayesian approach we focus on. A Bayesian analysis is of particular interest, both because suggested solutions such as shrinkage estimators and ridge regression can be interpreted in Bayesian terms and because, as Leamer (1978) notes, Bayesian estimators can be interpreted in terms of pooling two samples of data as Tobin (1950) did by combining cross-section and time-series data.

One can distinguish three cases. First, when there is exact collinearity between regressors, their individual coefficients are not identified, but given an informative prior their Bayesian posterior means are well defined. Second, the correlation matrix between regressors may be ill-conditioned in small samples, but has full rank for all T, including the case where $T \to \infty$. Here a Bayesian approach can compensate for the ill conditioned correlation matrix in small samples, but the posterior means converge to the true values in large samples, so for large samples there is little to choose between Bayesian and frequentist approaches. We consider the Bayesian analysis of a third, intermediate, case where the correlation matrix is of full rank for a finite T, but converges to a rank deficient matrix as T goes to infinity. So in the case of two regressors the correlation between them tends to ± 1 as $T \to \infty$. We call this the highly collinear case. Just as exact collinearity causes non-identification of the parameters, high collinearity can be viewed as weak identification of the parameters. This characterisation of the highly collinear case is in line with the notion of weak instruments and weak identification in the generalized method of moments, GMM, literature where the correlation of the instruments and the target variable is allowed to tend to zero with the sample size. See, for example, the survey by Stock, Wright, and Yogo (2002).

This representation allows us to examine the extent to which the Bayesian analysis is robust to the choice of prior. We analyse the asymptotic behaviour of the posterior mean and precision of the parameters of a linear regression model for exactly and highly collinear regressors, corresponding to the non-identified and weakly identified cases. Whereas in the identified case the posterior mean tends to its true value, in both the exactly collinear and highly collinear cases the posterior mean continues to depend on the priors even if $T \to \infty$, and the posterior precision increases at a rate slower than T. In the highly collinear case, the posterior means converge to normally distributed random variables whose mean and variance depend on prior means and prior precisions. The posterior distributions degenerate to fixed points in the polar cases of either exact collinearity or strong identification. This analysis also suggests diagnostics for the highly collinear case.

The analysis is related to Poirier (1998), Koop et al. (2013), Baumeister and Hamilton (2015), and Basturk et al. (2017); all of which consider Bayesian analysis of unidentified or weakly identified models. The focus in Koop et al. (2013) was on the behaviour of the posterior precision of the coefficient when the parameter was not identified or only weakly identified, here the focus will also be on the behaviour of the posterior mean.

Phillips (2016) provides a frequentist analysis of a similar case of near singular regressions for both least squares and instrumental variable estimators, and shows that in the case of asymptotically collinear regressors the estimators will be inconsistent and converge to random variables. We obtain similar asymptotic results for the Bayesian case. Cheng et al. (2017) comment that there is little discussion on the large sample behaviour of the posterior mean and examine asymptotic properties of posterior means obtained from simulations.

Many Bayesians emphasise finite T rather than asymptotic analysis. But we believe our asymptotic analysis is also relevant from a finite T perspective, since it addresses how data updates (changes in T) affect the posterior means and precisions. In the unidentified and weakly identified cases our analysis suggests that the posteriors remain dependent on the choice of the priors; and that this dependence does not diminish with successive Bayesian updates. It also follows that posterior mean of a weakly identified parameter (although well-defined for a finite T), will be much more sensitive to the choice of the priors as compared to the posterior mean of a strongly identified parameter. We illustrate these features with a Monte Carlo analysis.

The rest of the paper is organized as follows: Section 2 considers the exactly collinear case, where the parameters are not identified, to illustrate the influence of the priors on the posterior means and precisions as $T \to \infty$. Section 3 considers the highly collinear case, where the parameters are weakly identified. The strength of identification can be measured by a signal to noise ratio and Section 4 discusses the use of this ratio as a diagnostic indicator for collinearity. Section 5 contains a Monte Carlo Analysis to illustrate how the asymptotic results operate in finite samples. Section 6 uses the empirical relationship between stock returns and dividend yields to illustrate the application of this diagnostic. Section 7 contains some concluding comments. Some of the technical derivations are relegated to appendices. The computer code for the simulation exercise is available as an online supplement.

2 Exactly collinear regressors

This section examines the properties of the posterior means and precisions in the exactly collinear case as a benchmark for the highly collinear case. Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{u}$$

where $\mathbf{y}$ is a $T \times 1$ vector of observations on the dependent variable, $\mathbf{X}$ is a $T \times k$ matrix of observations on the k regressors, $\boldsymbol{\theta}$ a $k \times 1$ vector of unknown parameters and $\mathbf{u}$ is a $T \times 1$ vector of errors distributed independently of $\mathbf{X}$ as $N(\mathbf{0}, \sigma^2 \mathbf{I}_T)$. An element of $\boldsymbol{\theta}$, say θ_i is the parameter of interest and to simplify the exposition below we often assume that σ^2 is known. Since σ^2 does not appear in the expressions for the main results, this is not a strong assumption.

The least squares estimator is given by

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{y}$$

when $(\mathbf{X}'\mathbf{X})$ is non-singular. When $(\mathbf{X}'\mathbf{X})$ is rank deficient it may still be possible to estimate functions of $\boldsymbol{\theta}$ say $\boldsymbol{\beta} = \boldsymbol{b}'\boldsymbol{\theta}$.

However, even with exact collinearity, the Bayesian posterior distribution of $\boldsymbol{\theta}$ is well defined. Suppose that the prior distribution of $\boldsymbol{\theta}$ is $N(\underline{\boldsymbol{\theta}}, \underline{\mathbf{H}}^{-1})$, where $\underline{\boldsymbol{\theta}}$ is the prior mean and $\underline{\mathbf{H}}$ is the prior precision matrix of $\boldsymbol{\theta}$, which is a symmetric positive semi-definite matrix. Then based on a sample of T observations and known σ^2 the posterior mean of $\boldsymbol{\theta}$ is given by

$$\bar{\boldsymbol{\theta}}_T = \left(\sigma^{-2}T^{-1}\mathbf{X}'\mathbf{X} + T^{-1}\underline{\mathbf{H}}\right)^{-1} (\sigma^{-2}T^{-1}\mathbf{X}'\mathbf{y} + T^{-1}\underline{\mathbf{H}}\ \underline{\theta}), \tag{1}$$

and the covariance matrix of the posterior distribution of $\boldsymbol{\theta}$, denoted by $\bar{\mathbf{V}}$, is given by

$$\bar{\mathbf{V}} = \left(\sigma^{-2}\mathbf{X}'\mathbf{X} + \underline{\mathbf{H}}\right)^{-1}. \tag{2}$$

The posterior precision of θ_i, which we denote by $\bar{h}_{ii}$, is given by the inverse of the i^{th} diagonal element of $\bar{\mathbf{V}}$. We consider conjugate priors, which are widely used in a regression context, such as Bayesian VARs. This enables us to obtain analytical results and not have to resort to numerical methods.

When $T^{-1}\mathbf{X}'\mathbf{X}$ is non-singular for all $T > k$, then $\bar{\boldsymbol{\theta}}_T$ converges in probability to $\boldsymbol{\theta}^0$, as $T \to \infty$, where $\boldsymbol{\theta}^0$ is the true value of $\boldsymbol{\theta}$. But when there are exact linear dependencies amongst the regressors and $\mathbf{X}$ is rank deficient, the posterior mean remains well defined for finite T since $\left(\sigma^{-2}T^{-1}\mathbf{X}'\mathbf{X} + T^{-1}\underline{\mathbf{H}}\right)^{-1}$ exists even if $(\mathbf{X}'\mathbf{X})^{-1}$ does not. We consider below what happens to the posterior means (and precisions) as $T \to \infty$.

To simplify the exposition we consider the relatively simple case where $k = 2$ and the regression model is given by

$$y_t = \theta_1 x_{1t} + \theta_2 x_{2t} + u_t,\ u_t \sim IIDN(0, \sigma^2), \tag{3}$$

where the y_t and the regressors are measured as deviations from their means, and where $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ are the parameters of interest.

Suppose that there is exact collinearity of the form $x_{2t} = \phi x_{1t}$ for all t, and ϕ is a known non-zero constant. In this case

$$T^{-1}\mathbf{X}'\mathbf{X} = s_T^2 \kappa_\phi \kappa_\phi', \ \ T^{-1}\mathbf{X}'\mathbf{y} = s_T^2 \hat{\beta}_T \kappa_\phi \tag{4}$$

where $\hat{\beta}_T = s_{yT}/s_T^2$, $s_{yT} = T^{-1}\sum_{t=1}^{T} y_t x_{1t}$, $s_T^2 = T^{-1}\sum_{t=1}^{T} x_{1t}^2 > 0$, for all T, and $\kappa_\phi = (1, \phi)'$. Also note that the estimable function is

$$\hat{\beta}_T \rightarrow_p \beta^0 = \theta_1^0 + \phi \ \theta_2^0. \tag{5}$$

In the case where x_{1t} and x_{2t} are perfectly correlated, θ_1^0 and θ_2^0 are not unique but defined by all values of θ_1 and θ_2 that lie on the line $\beta = \theta_1 + \phi \ \theta_2$, for all values of $\beta \in \mathcal{R}$.

2.1 Posterior means in the exactly collinear case

We consider the limiting properties of the posterior means in the two regressor case, (3). Using (4) in (1) and after some algebra we have

$$\bar{\boldsymbol{\theta}}_T = \left(\kappa_\phi \kappa_\phi' + T^{-1}\mathbf{A}\right)^{-1} \left(\hat{\beta}_T \kappa_\phi + T^{-1}\mathbf{b}\right),$$

where

$$\mathbf{A} = (a_{ij}) = \left(\sigma^2/s_T^2\right) \begin{pmatrix} \underline{h}_{11} & \underline{h}_{12} \\ \underline{h}_{12} & \underline{h}_{22} \end{pmatrix},$$

$$\mathbf{b} = (b_i) = \frac{\sigma^2}{s_T^2} \underline{\mathbf{H}} \ \underline{\theta} = \frac{\sigma^2}{s_T^2} \begin{pmatrix} \underline{h}_{11}\underline{\theta}_1 + \underline{h}_{12}\underline{\theta}_2 \\ \underline{h}_{12}\underline{\theta}_1 + \underline{h}_{22}\underline{\theta}_2 \end{pmatrix}.$$

Therefore,

$$\bar{\theta}_{1,T} = \frac{\hat{\beta}_T\left(a_{22} - \phi a_{12}\right) + \phi\left(\phi b_1 - b_2\right) + T^{-1}\left(b_1 a_{22} - b_2 a_{12}\right)}{a_{11}\phi^2 - 2\phi a_{12} + a_{22} + T^{-1}(a_{11}a_{22} - a_{12}^2)}, \tag{6}$$

$$\bar{\theta}_{2,T} = \frac{b_2 - \phi b_1 - \hat{\beta}_T\left(a_{12} - \phi a_{11}\right) + T^{-1}\left(b_2 a_{11} - b_1 a_{12}\right)}{a_{11}\phi^2 - 2\phi a_{12} + a_{22} + T^{-1}(a_{11}a_{22} - a_{12}^2)}. \tag{7}$$

These are exact results, but to investigate the probability limits of the posterior means we only need to consider the first order terms. The derivations are given in Appendix A1.

$$\bar{\theta}_{1,T} = \theta_1^0 + \frac{\left(\underline{h}_{11}\phi^2 - \phi\underline{h}_{12}\right)}{\underline{h}_{11}\phi^2 - 2\phi\underline{h}_{12} + \underline{h}_{22}}\left(\underline{\theta}_1 - \theta_1^0\right) - \frac{\left(\phi\underline{h}_{22} - \phi^2\underline{h}_{12}\right)}{\underline{h}_{11}\phi^2 - 2\phi\underline{h}_{12} + \underline{h}_{22}}\left(\underline{\theta}_2 - \theta_2^0\right) + O_p(T^{-1}), \tag{8}$$

and

$$\bar{\theta}_{2,T} = \theta_2^0 - \frac{\left(\phi\underline{h}_{11} - \underline{h}_{12}\right)}{\underline{h}_{11}\phi^2 - 2\phi\underline{h}_{12} + \underline{h}_{22}}\left(\underline{\theta}_1 - \theta_1^0\right) + \frac{\left(\underline{h}_{22} - \phi\underline{h}_{12}\right)}{\underline{h}_{11}\phi^2 - 2\phi\underline{h}_{12} + \underline{h}_{22}}\left(\underline{\theta}_2 - \theta_2^0\right) + O_p(T^{-1}), \tag{9}$$

In the case where $\underline{h}_{12} = 0$, the results simplify to

$$p\lim_{T\to\infty}\left(\bar{\theta}_{1,T}\right) = \theta_1^0 + \frac{\phi^2\underline{h}_{11}}{\underline{h}_{11}\phi^2 + \underline{h}_{22}}\left(\underline{\theta}_1 - \theta_1^0\right) - \frac{\phi\underline{h}_{22}}{\underline{h}_{11}\phi^2 + \underline{h}_{22}}\left(\underline{\theta}_2 - \theta_2^0\right),$$

$$p\lim_{T\to\infty}(\bar{\theta}_{2,T}) = \theta_2^0 - \frac{\phi \underline{\mathrm{h}}_{11}}{\underline{\mathrm{h}}_{11}\phi^2 + \underline{\mathrm{h}}_{22}}\left(\underline{\theta}_1 - \theta_1^0\right) + \frac{\underline{\mathrm{h}}_{22}}{\underline{\mathrm{h}}_{11}\phi^2 + \underline{\mathrm{h}}_{22}}\left(\underline{\theta}_2 - \theta_2^0\right),$$

which are not equal to their true values and highlight the role of the prior means and precisions of both coefficients in the determination of the asymptotic posterior means. In the case where the prior precisions are set to be the same across the parameters and $\underline{\mathrm{h}}_{12} = 0$, (often done in practice) we have

$$p\lim_{T\to\infty}\left(\bar{\theta}_{1,T}\right) = \theta_1^0 + \frac{\phi^2}{1+\phi^2}\left(\underline{\theta}_1 - \theta_1^0\right) - \frac{\phi}{1+\phi^2}\left(\underline{\theta}_2 - \theta_2^0\right), \quad (10)$$

$$p\lim_{T\to\infty}(\bar{\theta}_{2,T}) = \theta_2^0 - \frac{\phi}{1+\phi^2}\left(\underline{\theta}_1 - \theta_1^0\right) + \frac{1}{1+\phi^2}\left(\underline{\theta}_2 - \theta_2^0\right), \quad (11)$$

and the limit of posterior means do not depend on the prior precisions, but do depend on both prior means, even asymptotically.

2.2 Posterior precisions in the exactly collinear case

Using (2) and noting that $x_{2t} = \phi x_{1t}$ we have

$$\begin{aligned}\bar{\mathbf{V}} &= \left(T\tilde{s}_T^2\kappa_\phi\kappa_\phi' + \underline{\mathbf{H}}\right)^{-1} = \begin{pmatrix} T\tilde{s}_T^2 + \underline{\mathrm{h}}_{11} & T\tilde{s}_T^2\phi + \underline{\mathrm{h}}_{12} \\ T\tilde{s}_T^2\phi + \underline{\mathrm{h}}_{12} & T\tilde{s}_T^2 + \phi^2\underline{\mathrm{h}}_{22}\end{pmatrix}^{-1} \\ &= \frac{1}{\left(T\tilde{s}_T^2 + \underline{\mathrm{h}}_{11}\right)\left(T\tilde{s}_T^2 + \phi^2\underline{\mathrm{h}}_{22}\right) - \left(T\tilde{s}_T^2\phi + \underline{\mathrm{h}}_{12}\right)^2}\begin{pmatrix} T\tilde{s}_T^2 + \phi^2\underline{\mathrm{h}}_{22} & -T\tilde{s}_T^2\phi - \underline{\mathrm{h}}_{12} \\ -T\tilde{s}_T^2\phi - \underline{\mathrm{h}}_{12} & T\tilde{s}_T^2 + \underline{\mathrm{h}}_{11}\end{pmatrix},\end{aligned}$$

where $\tilde{s}_T^2 = s_T^2/\sigma^2$. The posterior precision of θ_1 is given by the inverse of the first element of $\bar{\mathbf{V}}$, namely

$$\bar{h}_{11} = \frac{\left(T\tilde{s}_T^2 + \underline{\mathrm{h}}_{11}\right)\left(T\tilde{s}_T^2 + \phi^2\underline{\mathrm{h}}_{22}\right) - \left(T\tilde{s}_T^2\phi + \underline{\mathrm{h}}_{12}\right)^2}{T\tilde{s}_T^2 + \phi^2\underline{\mathrm{h}}_{22}},$$

which gives the following result for the average precision of θ_1

$$T^{-1}\bar{h}_{11} = (\tilde{s}_T^2 + T^{-1}\underline{\mathrm{h}}_{11}) - \left(\phi\tilde{s}_T^2 + T^{-1}\underline{\mathrm{h}}_{12}\right)\left(\phi^2\tilde{s}_T^2 + T^{-1}\underline{\mathrm{h}}_{22}\right)^{-1}\left(\phi\tilde{s}_T^2 + T^{-1}\underline{\mathrm{h}}_{21}\right),$$

and after some algebra yields

$$T^{-1}\bar{h}_{11} = T^{-1}\tilde{s}_T^2\left\{\frac{(\underline{\mathrm{h}}_{22}/\tilde{s}_T^2) + (\underline{\mathrm{h}}_{11}/\tilde{s}_T^2)\phi^2 + (\underline{\mathrm{h}}_{11}/\tilde{s}_T^2)T^{-1}(\underline{\mathrm{h}}_{22}/\tilde{s}_T^2) - 2\phi\underline{\mathrm{h}}_{21}/\tilde{s}_T^2 - T^{-1}\left(\underline{\mathrm{h}}_{21}/\tilde{s}_T^2\right)^2}{\phi^2 + T^{-1}(\underline{\mathrm{h}}_{22}/\tilde{s}_T^2)}\right\}.$$

It now readily follows that $\lim_{T\to\infty} T^{-1}\bar{h}_{11} = 0$, namely for any choice of priors and finite values of $\tilde{s}_T^2$, the average precision of θ_1 will tend to zero when the regressors are exactly collinear. This result contrasts to the identified case where the average precision tends to a non-zero constant. It is also instructive to consider the special case when the priors of θ_1 and θ_2 are independent, namely $\underline{\mathrm{h}}_{12} = \underline{\mathrm{h}}_{21} = 0$. In this case the above expression simplifies to

$$\bar{h}_{11} = \frac{\underline{\mathrm{h}}_{22} + \phi^2\underline{\mathrm{h}}_{11} + T^{-1}\underline{\mathrm{h}}_{11}\underline{\mathrm{h}}_{22}/\tilde{s}_T^2}{\phi^2 + T^{-1}(\underline{\mathrm{h}}_{22}/\tilde{s}_T^2)}.$$

Hence, the posterior precision ($\bar{h}_{11}$) of the unidentified parameter, θ_1, differs from its prior precision ($\underline{h}_{11}$) for all T, and as $T \to \infty$, even though θ_1 and θ_2 are assumed to be a priori independent. Also, for T sufficiently large we have

$$\lim_{T\to\infty} \bar{h}_{11} = \underline{h}_{11} + \phi^{-2}\underline{h}_{22},$$

which shows that the posterior precision is bounded in T, in contrast to the posterior precision of an identified parameter that rises linearly with T.

The extent to which the posterior precision deviates from the prior precision is determined by $\underline{h}_{22}/\phi^2$. It is also worth noting, however, that as T increases the posterior precision declines. This could be viewed as an indication that θ_1 is not identified. In the case where a parameter is identified we would expect the posterior precision to rise with T and eventually dominate the prior precision.

3 Highly collinear regressors

In practice, the case of exactly collinear regressors is only of pedagogical interest. In this section we investigate the role of the priors in regression analysis when the regressors are highly collinear and are expected to remain so even if we consider larger data sets. Following the literature on weak identification, we define the highly collinear case as being where the correlation matrix is full rank for a finite T, but tends to a rank deficient matrix as $T \to \infty$. Thus we model the collinearity of the regressors in (3) by

$$x_{2t} = \phi x_{1t} + \frac{\delta_T}{\sqrt{T}} v_t, \tag{12}$$

where v_t is a stationary process with zero means, distributed independently of x_{1t} and u_t such that

$$s_{vv,T} = T^{-1}\sum_{t=1}^{T} v_t^2 \to_p \sigma_v^2, \quad s_T^2 = T^{-1}\sum_{t=1}^{T} x_{1t}^2 \to_p \sigma_1^2, \tag{13}$$

$$T^{-1/2} s_T^{-2} \sum_{t=1}^{T} x_{1t} v_t \to_d N(0, \sigma_v^2), \quad T^{-1/2}\sum_{t=1}^{T} u_t v_t \to_d N(0, \sigma^2\sigma_v^2). \tag{14}$$

The coefficient δ_T in (12) controls the degree of collinearity between the two regressors. It is clear that the correlation between x_{1t} and x_{2t} is not perfect when T is finite, but when δ_T is constant, it tends to unity as $T \to \infty$. More specifically, denoting the correlation coefficient of x_{1t} and x_{2t} by ρ_T, we have

$$\rho_T = \frac{\phi + \frac{\delta_T}{\sqrt{T}}\left(\frac{T^{-1/2}\sum_{t=1}^{T} x_{1t}v_t}{s_T^2}\right)}{\sqrt{\phi^2 + 2\phi\frac{\delta_T}{\sqrt{T}}\left(\frac{T^{-1/2}\sum_{t=1}^{T} x_{1t}v_t}{s_T^2}\right) + \frac{\delta_T^2}{T}\left(\frac{s_{vv,T}}{s_T^2}\right)}},$$

which in view of (13) and (14) yields

$$\rho_T = \left(\frac{\phi}{|\phi|}\right)\left[1 + O_p\left(\frac{\delta_T}{\sqrt{T}}\right)\right]. \tag{15}$$

In finite samples ρ_T could take any value over the range $(-1, 1)$, but tends to ± 1, as $T \to \infty$. It tends to 1 if $\phi > 0$, and to -1 if $\phi < 0$. The above result can also be written equivalently as

$$\rho_T^2 = 1 + O_p\left(\frac{\delta_T}{\sqrt{T}}\right).$$

There is a one-to-one relationship between the degree of correlation of x_{1t} and x_{2t} and the degree of identifiability of θ_1 and θ_2. The different cases can be characterized in terms of δ_T. In the perfectly collinear case $\delta_T = 0$, for all T, and in the highly collinear case of weak identification δ_T is bounded in T. Strong identification requires $\delta_T^2 = \circleddash(T)$ where $\circleddash(T)$ denotes that δ_T^2 rises at the *same* rate as T, such that $\rho_T^2 < 1$, for all values of T, including as $T \to \infty$. The notation $f = \circleddash(T)$ differs from the standard big O notation, $f = O(T)$. The latter provides an upper bound on the expansion rate of the function in terms of T, whilst the former refers to the exact rate at which the function rises with T.

As noted above, this formulation is akin to the treatment of weak identification employed in the GMM literature. Where we have $\rho_T^2 \to 1$, as $T \to \infty$, in that literature a reduced form coefficient goes to zero as $T \to \infty$. For instance, Staiger and Stock (1997) consider the case of a single right hand side endogenous variable with reduced form coefficient π and introduce weak instrument asymptotics as a local to zero alternative of the form $\pi = \delta/\sqrt{T}$, where δ is a constant and T is the sample size. In a specification that is even more similar to ours, Sanderson and Windmeijer (2016) examine the case where there are two right hand side endogenous variables and consider weak instrument asymptotics local to a rank reduction of one of the form

$$\boldsymbol{\pi}_1 = \alpha \boldsymbol{\pi}_2 + \frac{\boldsymbol{\delta}}{\sqrt{T}}, \tag{16}$$

where $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ are vectors of parameters in the two reduced form equations, δ is a vector of constants and T is the sample size. Where (16) has the relation between the reduced form parameters a deterministic functions of the sample size, (12) postulates a stochastic relation between the regressors such that their correlation coefficient, ρ_T, tends to unity at the rate of $\delta_T/\sqrt{T}$, which corresponds to the local parameterization used in the weak instrument literature.

3.1 Posterior mean in the highly collinear case

The posterior mean of θ_1, namely $\bar{\theta}_{1,T}$, is derived in Appendix A2 and is given by (33)

$$\bar{\theta}_{1,T} = \theta_1^0 + \frac{\phi\left(\underline{\mathrm{h}}_{11}\phi - \underline{\mathrm{h}}_{12}\right)}{\lambda_T^2 + \boldsymbol{\psi}'\underline{\mathbf{H}}\boldsymbol{\psi}}\left(\underline{\theta}_1 - \theta_1^0\right) - \frac{\phi\left(\underline{\mathrm{h}}_{22} - \phi\underline{\mathrm{h}}_{12}\right)}{\lambda_T^2 + \boldsymbol{\psi}'\underline{\mathbf{H}}\boldsymbol{\psi}}\left(\underline{\theta}_2 - \theta_2^0\right)$$
$$- \left(\frac{\beta^0\phi\lambda_T}{\lambda_T^2 + \boldsymbol{\psi}'\underline{\mathbf{H}}\boldsymbol{\psi}}\right)\left(T^{-1/2}\sum_{t=1}^{T}\frac{v_t u_t}{\sigma_v \sigma}\right) + O_p\left(T^{-1/2}\right).$$

where $\boldsymbol{\psi} = (\phi, -1)'$, $\underline{\mathbf{H}} = \left(\underline{\mathrm{h}}_{ij}\right)$, and $\lambda_T^2 = \delta_T^2\sigma_v^2/\sigma^2$ is a signal-noise ratio that provides a summary measure of the relative importance of the collinearity for the analysis of the posterior mean. The above result generalizes equation (8), derived for the exactly collinear case, and reduces to it when $\delta_T = 0$.

Denoting the limit of δ_T as $T \to \infty$, by δ, (which could be 0 or ∞), then the posterior mean tends to a normal distribution that depends on prior means and precisions. More specifically we have

$$\bar{\theta}_{1,T} \to_d N\left(\mu, \omega^2\right), \text{ as } T \to \infty,$$

where

$$\mu = \theta_1^0 + \frac{\phi\left(\underline{\mathrm{h}}_{11}\phi - \underline{\mathrm{h}}_{12}\right)}{\lambda^2 + \psi'\underline{\mathbf{H}}\psi}\left(\underline{\theta}_1 - \theta_1^0\right) - \frac{\phi\left(\underline{\mathrm{h}}_{22} - \phi\underline{\mathrm{h}}_{12}\right)}{\lambda^2 + \psi'\underline{\mathbf{H}}\psi}\left(\underline{\theta}_2 - \theta_2^0\right),$$

and

$$\omega^2 = \frac{\left(\beta^0\phi\right)^2\lambda^2}{\left(\lambda^2 + \psi'\underline{\mathbf{H}}\psi\right)^2}.$$

The frequentist results in Phillips (2016, Theorem 1) match the above result that the posterior means do not converge to their true values and are normally distributed random variables, and show the similarity between classical and Bayesian approaches for weakly identified cases.

The nature of the limiting property of the posterior mean, $\bar{\theta}_{1,T}$, critically depends on the (population) signal-to-noise ratio $\lambda^2 = \delta^2\sigma_v^2/\sigma^2$. The signal, $\delta^2\sigma_v^2$, measures the extent to which x_{1t} and x_{2t} have "independent" variation in the regression of x_{2t} on x_{1t}, (12), while σ^2 is the measure of the noise in the regression. As will be discussed below this provides a measure of the strength of identification. The distribution of $\bar{\theta}_{1,T}$ degenerates to a fixed value only under the two polar cases of exact collinearity and strong identification. In the case of exact collinearity $\delta = \lambda = 0$, and we have $\omega^2 = 0$, and μ is the limit (as $T \to \infty$) of the posterior mean of θ_1 in the exactly collinear case discussed in Section 2.1. In the case where the parameters are strongly identified, $\delta_T^2 = \ominus(T)$, such that $\delta_T^2/T \to c > 0$, then $\omega^2 \to 0$, and $\mu \to \theta_1^0$.

3.2 Posterior precision in the highly collinear case

Turning to posterior precisions, using (2) we have

$$\bar{\mathbf{V}}^{-1} = T\tilde{s}_T^2\begin{pmatrix} 1 & \phi \\ \phi & \phi^2 \end{pmatrix} + \begin{pmatrix} \underline{\mathrm{h}}_{11} & \underline{\mathrm{h}}_{12} + \delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right) \\ \underline{\mathrm{h}}_{12} + \delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right) & \underline{\mathrm{h}}_{22} + \lambda_{21,T}^2 + 2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right) \end{pmatrix}, \tag{17}$$

where as before $\tilde{s}_T^2 = s_T^2/\sigma^2$, and

$$s_{1v,T} = T^{-1}\sum_{t=1}^{T} x_{1t}v_t,\; s_{vv,T} = T^{-1}\sum_{t=1}^{T} v_t^2,\; \lambda_{21,T}^2 = \delta_T^2\left(s_{vv,T}/\sigma^2\right).$$

The posterior precision of θ_1 is given by the inverse of the first element of $\bar{\mathbf{V}}$. The derivations are given in Appendix A3, where it is shown that,

$$\bar{h}_{11,T} = \frac{\tilde{s}_T^2\left(\underline{\mathrm{h}}_{11}\phi^2 + \lambda_T^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22}\right)}{\phi^2\tilde{s}_T^2 + 2\chi_T\phi T^{-1}z_T + T^{-1}\underline{\mathrm{h}}_{22} + T^{-1}\lambda_T^2} + \frac{-T^{-1}\chi_T^2 z_T^2 + 2\chi_T\left(\underline{\mathrm{h}}_{11}\phi - \underline{\mathrm{h}}_{12}\right)T^{-1}z_T}{\phi^2\tilde{s}_T^2 + 2\chi_T\phi T^{-1}z_T + T^{-1}\underline{\mathrm{h}}_{22} + T^{-1}\lambda_T^2}$$
$$\frac{+\underline{\mathrm{h}}_{11}T^{-1}\lambda_T^2 + T^{-1}\underline{\mathrm{h}}_{11}\underline{\mathrm{h}}_{22} - T^{-1}\underline{\mathrm{h}}_{12}^2}{\phi^2\tilde{s}_T^2 + 2\chi_T\phi T^{-1}z_T + T^{-1}\underline{\mathrm{h}}_{22} + T^{-1}\lambda_T^2}, \tag{18}$$

where $\chi_T = \delta_T \sigma_v \sigma_{x_1}/\sigma^2$,

$$z_T = \frac{T^{1/2} s_{1v,T}}{\sigma_{x_1}\sigma_v} = T^{-1/2}\sum_{t=1}^{T}\frac{x_{1t}v_t}{\sigma_{x_1}\sigma_v} \rightarrow_d N(0,1).$$

Hence, for a finite T the posterior precision of θ_1 is a nonlinear function of the random variable z_T, and itself is also a random variable. The limiting properties of $\bar{h}_{11,T}$, crucially depends on the limiting properties of δ_T (see (12)) as $T \rightarrow \infty$. In the highly collinear case, δ_T is bounded in T and we have

$$p\lim_{T\rightarrow\infty} \bar{h}_{11,T} = \frac{\left(\lambda^2 + \underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22}\right)}{\phi^2} = \frac{\lambda^2 + \psi'\underline{\mathbf{H}}\psi}{\phi^2},$$

where as before $\lambda^2 = \delta^2\sigma_v^2/\sigma^2 = p\lim_{T\rightarrow\infty}\delta_T^2\left(s_{vv,T}/\sigma^2\right)$. Similarly,

$$p\lim_{T\rightarrow\infty} \bar{h}_{22,T} = \lambda^2 + \phi^2\underline{\mathrm{h}}_{11} - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} = \lambda^2 + \psi'\underline{\mathbf{H}}\psi.$$

Hence, in the highly collinear case (where θ_1 and θ_2 are weakly identified), the posterior precision tends to a finite limit, which is qualitatively the same conclusion obtained for the exactly collinear case. Finally, in the strongly identified case, where $\delta_T^2/T \rightarrow c^2 > 0$, then $\lim_{T\rightarrow\infty}\left(T^{-1}\lambda_T^2\right) = c^2\sigma_v^2/\sigma^2$, and using this results in (18) we have

$$\begin{aligned} p\lim_{T\rightarrow\infty} T^{-1}\bar{h}_{11,T} &= \frac{\lim_{T\rightarrow\infty}\left(T^{-1}\lambda_T^2\right)}{\phi^2\sigma_{x_1}^2/\sigma^2 + \lim_{T\rightarrow\infty}\left(T^{-1}\lambda_T^2\right)} \\ &= \frac{c^2\sigma_v^2/\sigma^2}{\phi^2\sigma_{x_1}^2/\sigma^2 + c^2\sigma_v^2/\sigma^2} = \frac{c^2\sigma_v^2}{\phi^2\sigma_{x_1}^2 + c^2\sigma_v^2} > 0. \end{aligned}$$

Also using (12) it follows that $\phi^2\sigma_{x_1}^2 + c^2\sigma_v^2 = \sigma_{x_2}^2$, and hence in the strongly identified case

$$p\lim_{T\rightarrow\infty} T^{-1}\bar{h}_{11,T} = 1 - \rho^2,$$

where ρ is the population correlation coefficient between x_{1t} and x_{2t}. Therefore, as to be expected, in contrast to the highly collinear case, the posterior precision of strongly identified coefficients rise with T such that the average precision, $T^{-1}\bar{h}_{11,T}$, tends to a strictly positive constant. Also, as to be expected, the posterior precision does not depend on the priors when T is sufficiently large and the regression coefficients are strongly identified.

Finally, it is worth noting that the limiting property of the average precision is qualitatively the same irrespective of whether the parameters are not identified, the exactly collinear case, or weakly identified, the highly collinear case. In both cases the average precision tends to zero with T, although the rates at which this occurs does depend on whether the underlying parameter is weakly identified or not identified. This common feature does not extend to the posterior mean, whose limiting properties differ between the weakly identified and not identified cases.

4 Diagnostics for collinearity

As noted above, for large T the strength of identification is measured by the signal-to-noise ratio $\lambda^2 = \delta^2\sigma_v^2/\sigma^2$. The numerator, $\delta^2\sigma_v^2$, can be estimated from the OLS residuals of the regression of x_{2t} on x_{1t}, corresponding to (12), namely

$$\widehat{\delta^2\sigma_v^2} = \sum_{t=1}^{T} \left(x_{2t} - \hat{\phi}x_{1t}\right)^2.$$

The denominator, σ^2, can be estimated consistently from the regression of y_t on x_{1t} and x_{2t}, even if x_{1t} and x_{2t} are perfectly correlated, see Section 3.12 of Pesaran (2015). A consistent estimator of λ_T^2 is now given by:

$$\hat{\lambda}_T^2 = \frac{\widehat{\delta^2\sigma_v^2}}{\sigma^2} = \frac{\sum_{t=1}^{T}\left(x_{2t} - \hat{\phi}x_{1t}\right)^2}{T^{-1}\sum_{t=1}^{T}\left(y_t - \hat{\theta}_1 x_{1t} - \hat{\theta}_2 x_{2t}\right)^2}. \tag{19}$$

This collinearity diagnostic can also be written equivalently as

$$\hat{\lambda}_T^2 = \frac{T\hat{\sigma}_{2:1}^2}{\hat{\sigma}^2}, \tag{20}$$

where $\hat{\sigma}_{2:1}^2$ is the estimator of the error variance of the regression of x_{2t} on x_{1t}, and $\hat{\sigma}^2$ is the estimator of the error variance of the regression model. This will be zero in the case of exact collinearity.

We first consider the possibility of testing for weak identification and show that it is not feasible because of the presence of a nuisance parameter. The null hypothesis of weak identification of θ_1 or θ_2, can be written as

$$H_0: \ \delta_T^2 = c^2,$$

where c is a positive constant. The alternative hypothesis of strong identification is defined by

$$H_1: \ \delta_T^2 = \ominus(T).$$

Using (12), under the null hypothesis (and noting that all variables are measured as deviations from their means) we have

$$T\hat{\sigma}_{2:1}^2 = \mathbf{x}_2'\mathbf{M}_1\mathbf{x}_2 = c^2\left(\frac{\mathbf{v}'\mathbf{M}_1\mathbf{v}}{T}\right),$$

and hence

$$\hat{\lambda}_T^2 = \frac{\left(\frac{c^2\sigma_v^2}{\sigma^2}\right)\left(\frac{\mathbf{v}'\mathbf{M}_1\mathbf{v}}{T\sigma_v^2}\right)}{\frac{\mathbf{u}'\mathbf{M}\mathbf{u}}{T\sigma^2}},$$

where $\mathbf{v} = (v_1, v_2, ..., v_T)'$, $\mathbf{u} = (u_1, u_2, ..., u_T)'$, $\mathbf{M}_1 = \mathbf{I}_T - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1$, $\mathbf{M} = \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$, $\mathbf{X}_1 = (\tau_T, \mathbf{x}_1)$, $\mathbf{X} = (\tau_T, \mathbf{x}_1, \mathbf{x}_2)$, and τ_T is a $T \times 1$ vector of ones. For T large and by the Slutsky Theorem

$$\hat{\lambda}_T^2 \overset{a}{\sim} \lambda^2\left(\frac{\mathbf{v}'\mathbf{M}_1\mathbf{v}}{T\sigma_v^2}\right),$$

where $\lambda^2 = \left(\frac{c^2\sigma_v^2}{\sigma^2}\right)$, and $\hat{\lambda}_T^2 \to_p \lambda^2$. Consider now the standardized test statistic

$$\Delta_T' = \sqrt{\frac{T-2}{2}}\left[\left(\frac{T}{T-2}\frac{\hat{\lambda}_T^2}{\lambda^2} - 1\right)\right], \tag{21}$$

and suppose that v_t is $IIDN(0, \sigma_v^2)$. Then, since $\mathbf{M}_1$ is an idempotent matrix of rank $T-2$, we have

$$\Delta_T' = \frac{\sigma_v^{-2}\mathbf{v}'\mathbf{M}_1\mathbf{v} - (T-2)}{\sqrt{2\,(T-2)}} = \frac{\sum_{i=1}^{T-2}(\xi_i^2 - 1)/\sqrt{2}}{\sqrt{(T-2)}},$$

where ξ_i^2 are $IID(1,2)$. Hence, under H_0, $\Delta_T' \to_d N(0,1)$. An asymptotically equivalent version of Δ_T' is

$$\Delta_T = \sqrt{\frac{T}{2}}\left(\frac{\hat{\lambda}_T^2}{\lambda^2} - 1\right) \to_d N(0,1), \text{ under } H_0 \text{ and as } T \to \infty. \tag{22}$$

In practice, the implementation of the test is complicated by the fact that Δ_T depends on the nuisance constant λ^2. The test could only be implemented if one had a prior view about the value of λ.

Given that testing is not feasible because of the dependence of Δ_T on λ^2, an alternative strategy is to use $\hat{\lambda}_T^2$ as an indicator of high collinearity, with low values interpreted as evidence of weak identification of θ_1 (or θ_2). Under exact collinearity, $\hat{\lambda}_T^2 = 0$, and it might be expected to be close to zero in the highly collinear case. If identification is strong we would expect $\hat{\lambda}_T^2$ to rise with T. But if identification is weak, in the sense defined above, we would not expect $\hat{\lambda}_T^2$ to rise with T. Accordingly, collinearity is likely to be a problem if $\hat{\lambda}_T^2$ is small and does not increase much as T increases. This suggests estimating $\hat{\lambda}_T^2$ using expanding observation windows starting with the first T_0 observations and then plotting $\hat{\lambda}_\tau^2$, for $\tau = T_0, T_0+1, ..., T$ and check the rate at which $\hat{\lambda}_\tau^2$ rises with τ. Equivalently, one could consider whether $\tau^{-1}\hat{\lambda}_\tau^2$ remains bounded away from zero as τ is increased.

A scaled version of the high collinearity diagnostic statistic, $\hat{\lambda}_T^2$, is also related to the R^2 rule of thumb due to Klein (1962, p101) that considers multicollinearity is likely to be a problem if $R_{12}^2 > R_y^2$, where R_{12}^2 $(= R_{21}^2)$ is the squared correlation coefficient of x_{1t} and x_{2t}, and R_y^2 is the multiple correlation coefficient of the regression model, since.

$$\left(\frac{Var(y)}{Var(x_1)}\right)\hat{\lambda}_T^2 = T\left(\frac{1-R_{12}^2}{1-R_y^2}\right).$$

The above results and the diagnostic given by (20) generalize to regression models with more than two regressors. In the case of a linear regression model with k regressors (not counting the intercept) the high collinearity diagnostic statistic for the i^{th} regressors is given by

$$\hat{\lambda}_{iT}^2 = \frac{T\hat{\sigma}_i^2}{\hat{\sigma}^2}, \text{ for } i = 1, 2, ..., k, \tag{23}$$

where $\hat{\sigma}_i^2$ is the estimator of the error variance of the regression of the i^{th} regressor on the remaining regressors, and $\hat{\sigma}^2$ is the estimator of the underlying regression model. Once again expanding window estimates of $T^{-1}\hat{\lambda}_{iT}^2$ can provide useful indication of the weak identification of the i^{th} coefficient in the regression model. There would be a collinearity problem if $\hat{\lambda}_{i\tau}^2$ for $\tau = T_0, T_0+1, ..., T$ do not exhibit an upward trend as the window size is increased. The relative size of this measure for different regressors also indicates their relative sensitivity to collinearity.

In cases where T is short one could follow Koop et al. (2013), and consider estimates of $T^{-1}\hat{\lambda}_{i,T}^2$ using bootstrapped samples generated using the regression model and the marginal regressions of x_{it} on the remaining regressors.

5 Monte Carlo Analysis

We conduct a number of Monte Carlo experiments to investigate the extent to which our asymptotic theoretical results apply in finite samples. We consider a regression with two serially correlated and multicollinear regressors x_{1t} and x_{2t}. Given the Monte Carlo design and parameter values chosen for δ_T, that controls the correlation between the regressors, we consider how the posterior means and precisions of the regression coefficients evolve as T increases.

5.1 Design

For replications $r = 1, 2, ..., 2000$, we generate x_{1t} as

$$x_{1t} = \gamma x_{1,t-1} + \sqrt{1-\gamma^2}\varepsilon_t, \quad \varepsilon_t \sim IIDN(0,1),$$

for $t = -49, -48, ..., 0, 1, 2, ...T$, with $x_{1,-50} = 0$. We drop the first 50 observations to reduce the impact of the initial observation on x_{1t} and use x_{1t}, $t = 1, 2, ..., T$ in the simulations. Unconditionally $x_{1t} \sim N(0,1)$. We generate x_{2t} as

$$x_{2t} = \phi x_{1t} + \left(\frac{\delta_T}{\sqrt{T}}\right) v_t, \; v_t \sim IIDN(0,1),$$

so that $x_{2t} \sim N(0, \phi^2 + T^{-1}\delta_T^2)$, and $Cov\,(x_{1t}, x_{2t}) = \phi$. We also note that x_{2t} follows a first order moving average process which reduces to an AR(1) process under the highly collinear case where $T^{-1}\delta_T^2 \to 0$. We generate y_t as

$$y_t = \theta_1 x_{1t} + \theta_2 x_{2t} + u_t, \; u_t \sim IIDN(0,1),$$

We fix $\gamma = 0.9$; $\theta_1 = \theta_2 = 1$, $\phi = 2$ and consider the following values of δ_T:

$$\delta_T = 0: \text{ exactly collinear,}$$
$$\delta_T = 1 \text{ and } 5: \text{ highly collinear,}$$
$$\delta_T = T^{1/2}: \text{ not highly collinear.}$$

The priors for means and precisions are set as

$$\underline{\boldsymbol{\theta}} = \begin{pmatrix} \underline{\theta}_1 \\ \underline{\theta}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\underline{\mathbf{H}} = \begin{pmatrix} \underline{h}_{11} & \underline{h}_{12} \\ \underline{h}_{21} & \underline{h}_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

For each δ_T and for each replication we compute equations (1) and (2) above, repeated here for convenience:

$$\begin{pmatrix} \bar{\theta}_{1T} \\ \bar{\theta}_{2T} \end{pmatrix} = \bar{\boldsymbol{\theta}}_T = \left(\sigma^{-2}T^{-1}\mathbf{X}'\mathbf{X} + T^{-1}\underline{\mathbf{H}}\right)^{-1}\left(\sigma^{-2}T^{-1}\mathbf{X}'\mathbf{y} + T^{-1}\underline{\mathbf{H}}\ \underline{\theta}\right), \tag{24}$$

and

$$\bar{\mathbf{V}} = \left(\sigma^{-2}\mathbf{X}'\mathbf{X} + \underline{\mathbf{H}}\right)^{-1}, \tag{25}$$

where

$$\underset{T\times 2}{\mathbf{X}} = \begin{pmatrix} \underset{T\times 1}{\boldsymbol{x}_1} & \underset{T\times 1}{\boldsymbol{x}_2} \end{pmatrix},$$

and

$$\boldsymbol{x}_1 = \begin{pmatrix} x_{11} \\ x_{12} \\ \ldots \\ x_{1T} \end{pmatrix},\ \boldsymbol{x}_2 = \begin{pmatrix} x_{21} \\ x_{22} \\ \ldots \\ x_{2T} \end{pmatrix}.$$

As in the theoretical derivations we treat σ^2 as known and set $\sigma^2 = 1$ when calculating (24) and (25).

5.2 Expected outcomes

In the not highly collinear case where $\delta_T = T^{1/2}$,

$$E(x_{1t}^2) = 1 \text{ and } E(x_2^2) = 1 + \phi^2,\ Cov(x_{1t}, x_{2t}) = \phi,$$

which gives the following population value for the correlation coefficient between x_{1t} and x_{2t} :

$$\rho_T = \frac{\phi}{\left(\phi^2 + T^{-1}\delta_T^2\right)^{1/2}} = \frac{\phi^2}{1+\phi^2},\text{ if } \delta_T = T^{1/2}.$$

Also note that in the not collinear case $p\lim_{T\to\infty} T^{-1}\bar{h}_{11,T} = 1 - \rho^2$.

For exactly collinear case $\delta = 0$ and when, as here, the prior precisions are the same across parameters and $\underline{\mathrm{h}}_{12} = 0$, we have using (10) and (11) above

$$p\lim_{T\to\infty}\left(\bar{\theta}_{1,T}\right) = \theta_1^0 + \frac{\phi^2}{1+\phi^2}\left(\underline{\theta}_1 - \theta_1^0\right) - \frac{\phi}{1+\phi^2}\left(\underline{\theta}_2 - \theta_2^0\right),$$
$$p\lim_{T\to\infty}(\bar{\theta}_{2,T}) = \theta_2^0 - \frac{\phi}{1+\phi^2}\left(\underline{\theta}_1 - \theta_1^0\right) + \frac{1}{1+\phi^2}\left(\underline{\theta}_2 - \theta_2^0\right),$$

For the choices $\underline{\theta}_1 = \underline{\theta}_2 = 0$, $\theta_1^0 = \theta_2^0 = 1$ and $\phi = 2$, we have

$$p\lim_{T\to\infty}\left(\bar{\theta}_{1,T}\right) = 1 - \frac{\phi^2}{1+\phi^2} + \frac{\phi}{1+\phi^2} = \frac{1+\phi}{1+\phi^2} = \frac{3}{5}, \tag{26}$$
$$p\lim_{T\to\infty}(\bar{\theta}_{2,T}) = 1 + \frac{\phi}{1+\phi^2} - \frac{1}{1+\phi^2} = \frac{\phi(\phi+1)}{1+\phi^2} = \frac{6}{5}. \tag{27}$$

In the strongly identified case where there is no collinearity problem and $\delta_T = T^{1/2}$, the regressors are still quite highly correlated for this value of ϕ, and the population value for the squared correlation coefficient between x_{1t} and x_{2t} is

$$\rho_T^2 = \frac{\phi^2}{1+\phi^2} = \frac{4}{5}.$$

In addition, for this case since $\sigma^2 = 1$, then

$$V(\widehat{\boldsymbol{\theta}}\ ) = \left(\mathbf{X}'\mathbf{X}\right)^{-1} = \begin{pmatrix} \phi^2 + 1 & -\phi \\ -\phi & 1 \end{pmatrix}.$$

Precision is the inverse of the diagonal elements of the variance covariance matrix so asymptotically

$$p\lim_{T\to\infty} T^{-1}\bar{h}_{11,T} = 1/(1+\phi^2) = \frac{1}{5} \tag{28}$$

$$p\lim_{T\to\infty} T^{-1}\bar{h}_{22,T} = 1 \tag{29}$$

5.3 Results

We first consider the distribution of the posterior means for $T = 1000$. Figure 1, shows the distribution of the posterior mean in the exactly collinear case where $\delta = 0$. The distribution is tightly clustered around the values of $\bar{\theta}_{1,T} = 0.6$ and $\bar{\theta}_{2,T} = 1.2$ as expected from (26) and (27). The fact that the posteriors differ from the priors may be taken to indicate that there is learning about the true values, but this is not the case. The posterior means are just functions of ϕ, the correlation between x_{1t} and x_{2t}, not the true values of θ_1 and θ_2.

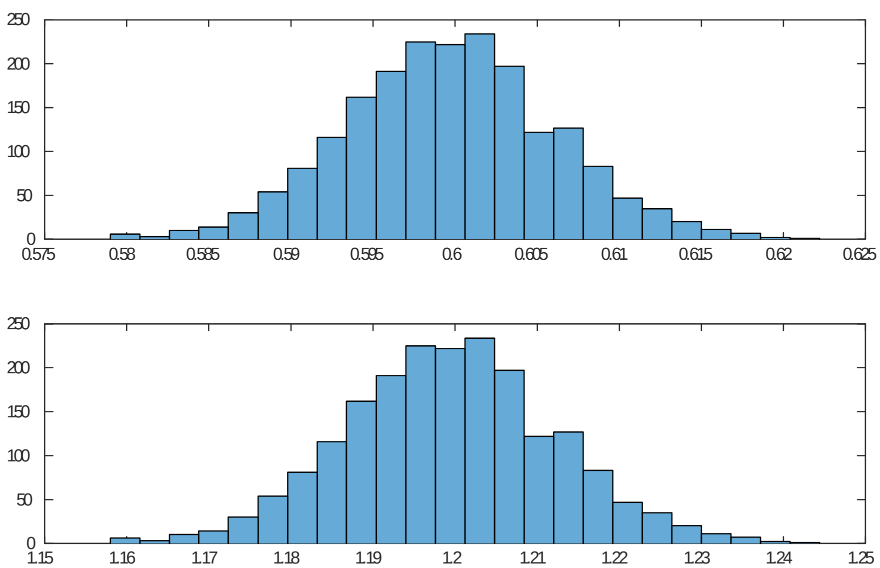

Figure 1: Distributions of posterior means for θ_1 (upper) and θ_2 (lower) for $T = 1000$ and $\delta = 0$

Figure 2 shows the distribution of the posterior means for a highly collinear case where $\delta = 1$. The posterior means are distributed around the same values of 0.6 and 1.2 as expected, though the distributions are much more dispersed..

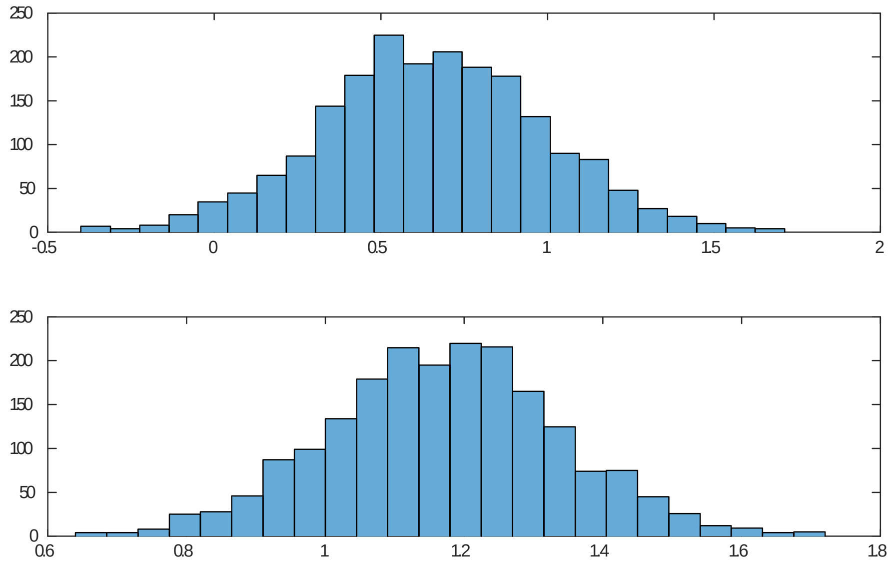

Figure 2: Distributions of posterior means for θ_1 (upper) and θ_2 (lower) for $T = 1000$ and $\delta = 1$

Figure 3 shows the distribution of the posterior mean for the case where the regressors are not collinear and $\delta = T^{1/2}$. As expected, the posterior means are distributed around true values of the parameters $\theta_1 = \theta_2 = 1$.

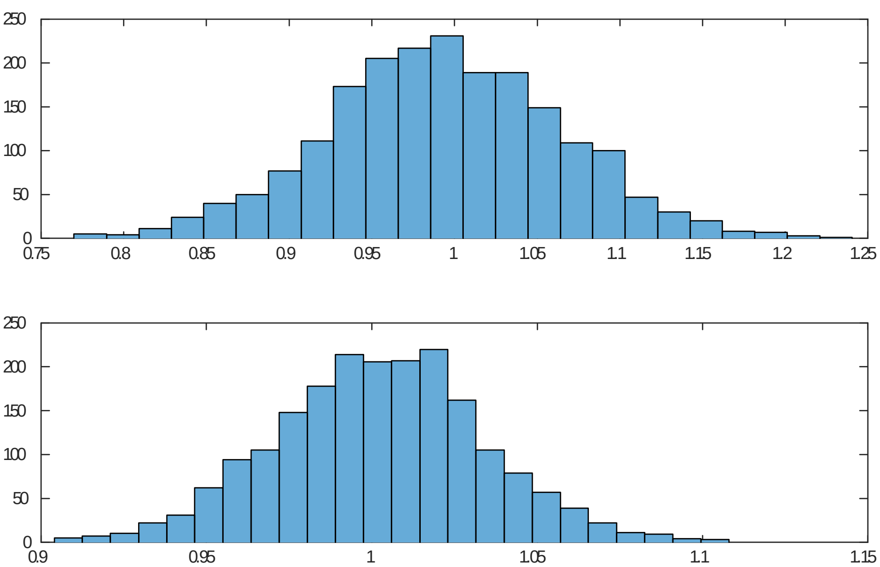

Figure 3: Distributions of posterior means for θ_1 (upper) and θ_2 (lower) for $T = 1000$ and $\delta = T^{1/2}$

We next examine the behaviour of the posterior precisions as the sample size changes. When the regressors are not collinear, the posterior precision of each coefficient should rise with T, so the average precision should go to a constant. When the regressors are exactly or highly collinear the posterior precision does not rise with T and the average precision goes to zero as T goes to infinity. Figure 4 plots the values of $T^{-1}\bar{h}_{11,T}$ and Figure 5 of $T^{-1}\bar{h}_{22}$ against $T = T_0, T_0+1, ..., 1000$, where $T_0 = 20$. Four different values of δ_T are shown on the same graph. The values are the exactly collinear case $\delta = 0$, two highly collinear cases $\delta = 1$ and $\delta = 5$, and the not highly collinear case, $\delta = T^{1/2}$. We are interested in how fast the posterior precisions of the exactly and highly collinear cases go to zero. The simulations match the theoretical results and show the asymptotic properties are important for sample sizes that occur in practice. For the not collinear case the average precisions have converged to their theoretical values given by (28) (29) by $T = 200$. For the exactly collinear and highly collinear cases the average precisions go to zero. For $\delta = 0$ and $\delta = 1$ they are close to zero by $T = 200$, for $\delta = 5$ by $T = 1000$.

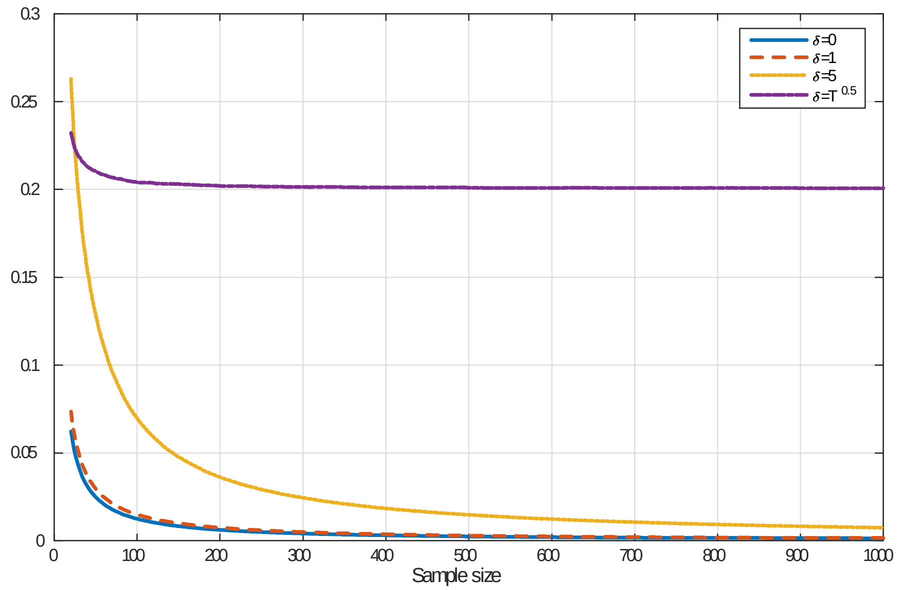

Figure 4: Simulated average posterior precision of θ_1

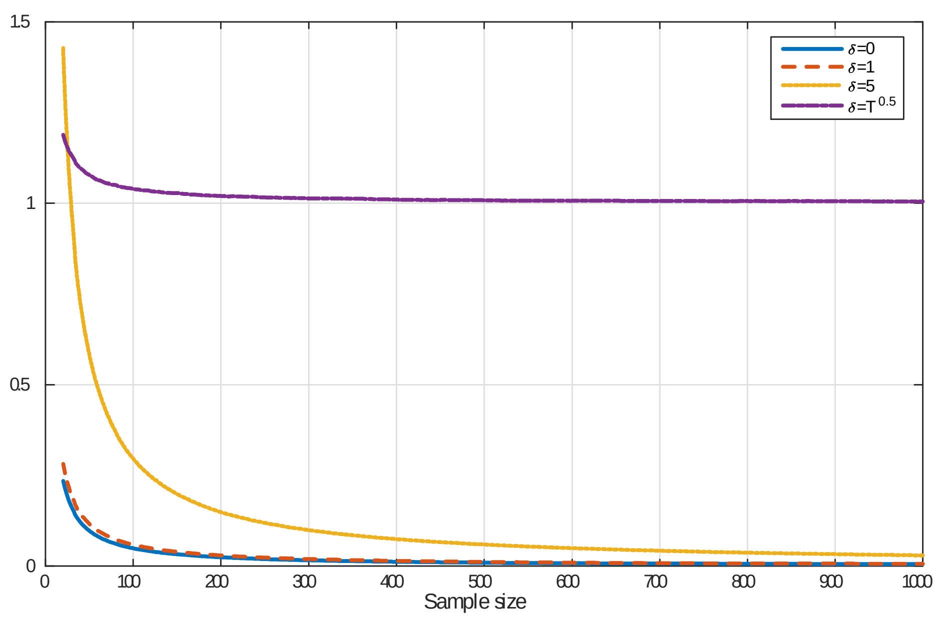

Figure 5: Simulated average posterior precision of θ_2

The Monte Carlo simulations show that the asymptotic results are relevant for sample sizes that are likely to be encountered in practice.

6 An empirical illustration

We use the example of predicting excess stock returns by the dividend yield. Stambaugh (1999) prompted a large literature on predictive regressions by showing that in regressions of rates of return on lagged stochastic regressors, such as dividend yields, the OLS estimator's finite-sample properties can depart substantially from the standard regression setting. He also showed that the Bayesian posterior distributions for the regression parameters are sensitive to prior beliefs about the autocorrelation of the regressor and whether the initial observation of the regressor is specified as fixed or stochastic.

We use Robert Shiller's online (http://www.econ.yale.edu/~shiller/data.htm) monthly data over the period 1871m1 2017m8. Monthly real excess returns on Standard & Poor 500 (SP500), denoted by y_t, are computed as

$$y_t = \left(\frac{s_t - s_{t-1}}{s_{t-1}}\right) + \frac{d_t}{s_{t-1}} - r_{t-1},$$

where $s_t = SP500_t/CPI_t$, $d_t = DIV_t/(12 * CPI_t)$, $SP500_t$ is the SP500 price index, CPI_t is the consumer price index, DIV_t is the annual rate of dividends paid on SP500, and r_t is the real return on ten year US government bond computed as

$$r_t = \left[(1 + GS10_t/100)^{1/12} - 1\right] - \pi_t,$$

where $GS10_t$ is the 10-Year Treasury Constant Maturity Rate per annum, and π_t is the rate of inflation computed as $\pi_t = (CPI_t - CPI_{t-1})/CPI_{t-1}$. The dividend yield variable is defined by $x_t = \ln(d_t/s_t)$. We consider the predictive regressions

$$y_t = \alpha_y + \lambda_y y_{t-1} + \theta_1 x_{1t} + \theta_2 x_{2t} + u_t, \tag{30}$$

where $x_{it} = x_{t-i}$, for $i = 1, 2$, and compute recursive estimates of $\sigma^2 = Var(u_t)$ using expanding windows starting with $1872m1$ and ending at $2017m8$, namely 1746 monthly observations. We denote these recursive estimates by $\hat{\sigma}^2_\tau$. We also consider the recursive estimates of the following auxiliary regression

$$x_{1t} = \alpha_x + \phi x_{2t} + \lambda_x y_{t-1} + v_t, \tag{31}$$

and compute the recursive estimates of $\sigma_1^2 = Var(v_t)$, which we denote by $\hat{\sigma}^2_{1,\tau}$. The recursive estimates of the collinearity indicator of θ_1 is now given by

$$\tau^{-1}\hat{\lambda}^2_{1,\tau} = \frac{\hat{\sigma}^2_{1,\tau}}{\hat{\sigma}^2_\tau}.$$

In the case where θ_1 is strongly identified we would expect $\hat{\lambda}^2_{1,\tau}$ to rise *linearly* with τ, or equivalently that $\tau^{-1}\hat{\lambda}^2_{1,\tau}$ to remain reasonably constant over the period $1872m1 - 2017m8$. To avoid the large sample variations when τ is small we drop the first 100 observations and show the values of $\tau^{-1}\hat{\lambda}^2_{1,\tau}$ over the period $\tau = 1880m1 - 2017m8$ in Figure 6.

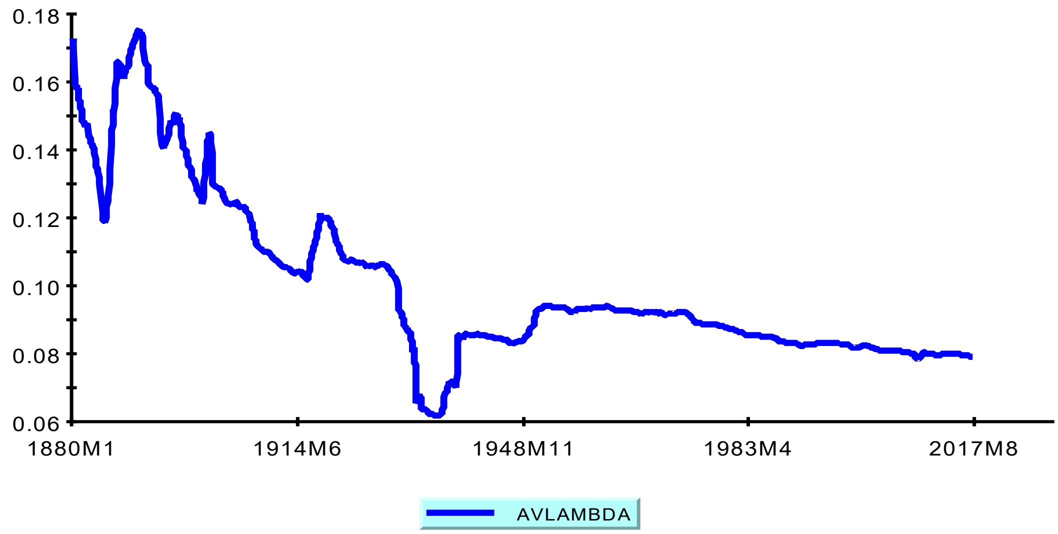

Figure 6: The recursive estimates of $\tau^{-1}\hat{\lambda}_{1\tau}^2$ for the dividend yield variable, x_{1t}

As can be seen, the high collinearity indicator has been falling over the sample with the exception of a brief period after the stock market crash of 1929. This suggests that the coefficients of the dividend yield variables are likely to be weakly identified. To illustrate the effect on the coefficients we plot the recursive coefficient of the lagged dividend yield variable and its two standard error band in Figure 7. As can be seen the error band covers zero over the whole period and the standard error does not reduce with the expanding sample. Adding more data does not seem effective in resolving the multicollinearity problem.

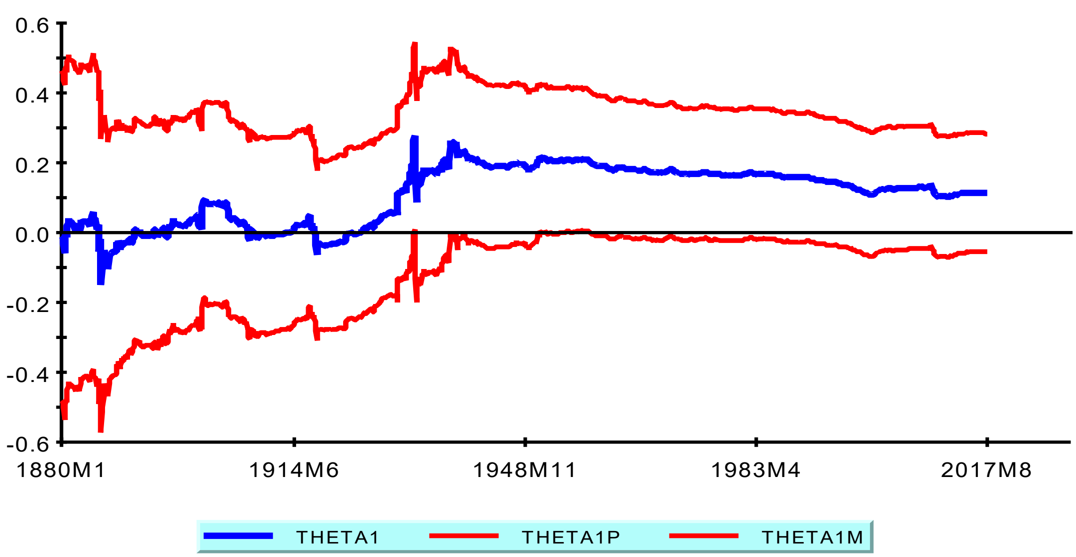

Fig 7: Recursive estimates of the coefficient of x_{1t} and its two standard error bands

To highlight the effect of the sample size on the variance of the estimated coefficients, we plot T times the estimated variance of the coefficient of the lagged dividend yield in Figure 8. In the strongly identified case, the estimates of the variance should fall at the rate of T, so T times the variance should be more or less stable as the sample size is increased. As can be seen from the figure, the opposite seems to be true, confirming that the multicollinearity problem is not being solved with more data.

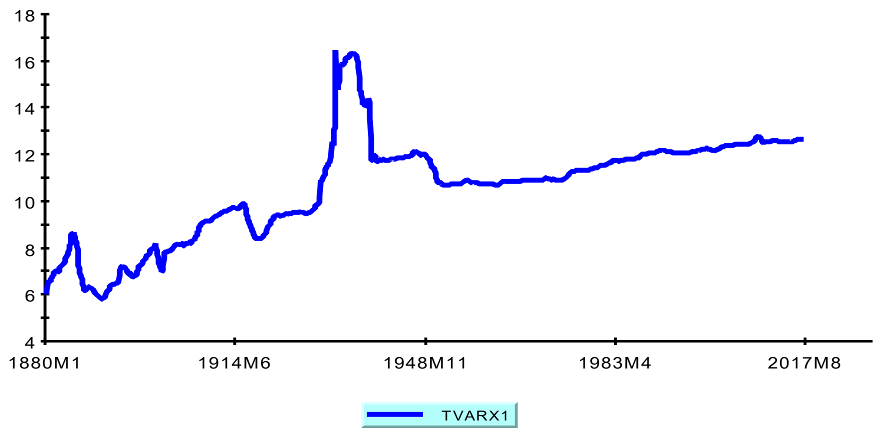

Figure 8: Plot of $T\ \widehat{Var}(\widehat{\theta}_1)$

Finally, we should emphasise that the theoretical analysis in this paper has been carried out assuming a regression model with stable coefficients. However, it is reasonable to expect that some of the observed time variations in the estimates could be due to parameter instability, particularly if we consider the estimates for the years following the stock market crash of 1929. The treatment of the multicollinearity problem in the presence of structural instability is a topic which falls outside the scope of the present paper and requires further research.

7 Conclusion

We have considered a Bayesian approach to collinearity among regressors. In the multicollinear case, where there are high but not perfect correlations, the coefficients are strongly identified and as the sample size gets large the Bayesian posterior mean converges to the true value of the parameter. In the exactly collinear case the posterior means converge to constants which depend on the priors and the posterior precision is bounded in T. In the highly collinear case where there are high correlations in finite samples and the data matrix becomes singular in the limit as $T \to \infty$, the posterior means converge to normally distributed random variables whose mean and variance depend on the priors for coefficients and precision. The distribution of this random variable degenerates to fixed points in the polar cases of either where the parameters are not identified, exact collinearity, or where the parameters are strongly identified. The analysis suggests an indicator of collinearity, $\hat{\lambda}^2_{i,T}$, a measure of the signal to noise ratio, for the *ith* regressor, which is zero in the exactly collinear case and rises with T in the strongly identified case. It is related to the R^2 rule of thumb due to Klein. We derive the distribution of this measure, which would allow it to be used as the basis for a test, except that it depends on a nuisance statistic. Thus it seems more useful as an estimated diagnostic for colinearity, since the size of $\hat{\lambda}_{i,T}^2$ and how it changes with T can be indicative of highly collinear relations.

Because the posterior mean can go to a random variable as the sample size increases in the highly collinear case of weak identification, it is not a reliable indicator. The posterior precision, which increases with T in the strongly identified case, provides a better indicator and our suggested diagnostic can be seen as a frequentist counterpart to the posterior precision.

Acknowledgements

A previous version of the paper was distributed as CESifo Working Paper 6785 under the title of "*Posterior Means and Precisions of the Coefficients in Linear Models with Highly Collinear Regressors*". We are grateful to Grigory Franguridi for carrying out the simulations and to Alex Chudik, Herman van Dijk, Gary Koop, Andrea Nocera, Adrian Pagan, Peter Phillips, Dale Poirier and two referees for useful comments on earlier versions.

Appendices

A1. Derivation of the probability limit for the posterior mean, $\bar{\boldsymbol{\theta}}_T$, in the exactly colinear case

First consider $\bar{\theta}_{1,T}$ given by (6):

$$\bar{\theta}_{1,T} = \frac{\hat{\beta}_T\left(\underline{h}_{22}-\phi\underline{h}_{12}\right)+\phi\left[\phi\underline{h}_{11}\underline{\theta}_1-\underline{h}_{12}\underline{\theta}_1+\phi\underline{h}_{12}\underline{\theta}_2-\underline{h}_{22}\underline{\theta}_2\right]}{\underline{h}_{11}\phi^2-2\phi\underline{h}_{12}+\underline{h}_{22}}+O_p(T^{-1}),$$
$$= \frac{\hat{\beta}_T\left(\underline{h}_{22}-\phi\underline{h}_{12}\right)+\phi\left(\phi\underline{h}_{11}-\underline{h}_{12}\right)\underline{\theta}_1+\phi\left(\phi\underline{h}_{12}-\underline{h}_{22}\right)\underline{\theta}_2}{\underline{h}_{11}\phi^2-2\phi\underline{h}_{12}+\underline{h}_{22}}+O_p(T^{-1}).$$

Then taking probability limits (noting that $\hat{\beta}_T\rightarrow_p\theta_1^0+\phi\;\theta_2^0$), we have

$$p\lim_{T\to\infty}\left(\bar{\theta}_{1,T}\right) = \frac{\theta_1^0\left(\underline{h}_{22}-\phi\underline{h}_{12}\right)+\phi\left(\phi\underline{h}_{11}-\underline{h}_{12}\right)\underline{\theta}_1+\theta_2^0\left(\phi\underline{h}_{22}-\phi^2\underline{h}_{12}\right)-\phi\left(\underline{h}_{22}-\phi\underline{h}_{12}\right)\underline{\theta}_2}{\underline{h}_{11}\phi^2-2\phi\underline{h}_{12}+\underline{h}_{22}},$$
$$= \frac{\left(\underline{h}_{22}-\phi\underline{h}_{12}\right)\theta_1^0+\phi\left(\phi\underline{h}_{11}-\underline{h}_{12}\right)\underline{\theta}_1+\phi\left(\underline{h}_{22}-\phi\underline{h}_{12}\right)\left(\theta_2^0-\underline{\theta}_2\right)}{\underline{h}_{11}\phi^2-2\phi\underline{h}_{12}+\underline{h}_{22}},$$
$$= \theta_1^0+\frac{\phi\left(\phi\underline{h}_{11}-\underline{h}_{12}\right)\left(\underline{\theta}_1-\theta_1^0\right)-\phi\left(\underline{h}_{22}-\phi\underline{h}_{12}\right)\left(\underline{\theta}_2-\theta_2^0\right)}{\underline{h}_{11}\phi^2-2\phi\underline{h}_{12}+\underline{h}_{22}}.$$

Similarly,

$$\bar{\theta}_{2,T} = \frac{\hat{\beta}_T\left(\phi\underline{h}_{11}-\underline{h}_{12}\right)+\left(\underline{h}_{12}-\phi\underline{h}_{11}\right)\underline{\theta}_1+\left(\underline{h}_{22}-\phi\underline{h}_{12}\right)\underline{\theta}_2}{\underline{h}_{11}\phi^2-2\phi\underline{h}_{12}+\underline{h}_{22}}+O_p(T^{-1}),$$

and

$$p\lim_{T\to\infty}\left(\bar{\theta}_{2,T}\right) = \frac{\left(\theta_1^0+\phi\theta_2^0\right)\left(\phi\underline{h}_{11}-\underline{h}_{12}\right)+\left(\underline{h}_{12}-\phi\underline{h}_{11}\right)\underline{\theta}_1+\left(\underline{h}_{22}-\phi\underline{h}_{12}\right)\underline{\theta}_2}{\underline{h}_{11}\phi^2-2\phi\underline{h}_{12}+\underline{h}_{22}},$$
$$= \frac{\phi\theta_2^0\left(\phi\underline{h}_{11}-\underline{h}_{12}\right)+\left(\underline{h}_{22}-\phi\underline{h}_{12}\right)\underline{\theta}_2+\left(\phi\underline{h}_{11}-\underline{h}_{12}\right)\left(\theta_1^0-\underline{\theta}_1\right)}{\underline{h}_{11}\phi^2-2\phi\underline{h}_{12}+\underline{h}_{22}},$$
$$= \theta_2^0+\frac{-\left(\phi\underline{h}_{11}-\underline{h}_{12}\right)\left(\underline{\theta}_1-\theta_1^0\right)+\left(\underline{h}_{22}-\phi\underline{h}_{12}\right)\left(\underline{\theta}_2-\theta_2^0\right)}{\underline{h}_{11}\phi^2-2\phi\underline{h}_{12}+\underline{h}_{22}}.$$

Let $\xi'=(-\phi,1)$, so $X\xi=\mathbf{0}$ then $\underline{h}_{11}\phi^2-2\phi\underline{h}_{12}+\underline{h}_{22}=\xi'\underline{\mathbf{H}}\xi$, and

$$p\lim_{T\to\infty}\left(\bar{\boldsymbol{\theta}}_T\right) = \boldsymbol{\theta}^0+\frac{1}{\underline{h}_{11}\phi^2-2\phi\underline{h}_{12}+\underline{h}_{22}}\begin{pmatrix}\phi^2 & -\phi\\ -\phi & 1\end{pmatrix}\begin{pmatrix}\underline{h}_{11} & \underline{h}_{12}\\ \underline{h}_{12} & \underline{h}_{22}\end{pmatrix}\left(\underline{\boldsymbol{\theta}}-\boldsymbol{\theta}^0\right)$$
$$= \boldsymbol{\theta}^0+\xi\left(\xi'\underline{\mathrm{H}}\xi\right)^{-1}\xi'\underline{\mathbf{H}}\left(\underline{\theta}-\theta^0\right).$$

Clearly, we have $p\lim_{T\to\infty}\left(\bar{\boldsymbol{\theta}}_T\right)=\boldsymbol{\theta}^0$, if $\underline{\boldsymbol{\theta}}=\boldsymbol{\theta}^0$, a sort of self-fulfillling belief.

Finally,

$$\bar{\theta}_{1,T}+\phi\bar{\theta}_{2,T} = \frac{\hat{\beta}_T\left(\phi^2a_{11}-2\phi a_{12}+a_{22}\right)+\frac{1}{T}\left(b_1a_{22}-b_2a_{12}\right)+\frac{1}{T}\phi\left(b_2a_{11}-b_1a_{12}\right)}{a_{11}\phi^2-2\phi a_{12}+a_{22}+T^{-1}\left[a_{11}a_{22}-a_{12}^2\right]}.$$

or

$$\bar{\theta}_{1,T}+\phi\bar{\theta}_{2,T} = \hat{\beta}_T+\frac{1}{T}\left[\frac{\left(b_1a_{22}-b_2a_{12}\right)+\phi\left(b_2a_{11}-b_1a_{12}\right)-\hat{\beta}_T\left[a_{11}a_{22}-a_{12}^2\right]}{a_{11}\phi^2-2\phi a_{12}+a_{22}}\right]+O(T^{-2})$$

Hence

$$p\lim_{T\to\infty}\left(\bar{\theta}_{1,T}+\phi\bar{\theta}_{2,T}\right)=p\lim_{T\to\infty}\left(\hat{\beta}_T\right)=\theta_1^0+\phi\theta_2^0.$$

Which is the only estimable function possible in a classical setting.

In the case where $\underline{\text{h}}_{12}=0$, the above results simplify to

$$p\lim_{T\to\infty}\left(\bar{\theta}_{1,T}\right)=\theta_1^0+\frac{\phi^2\underline{\text{h}}_{11}}{\underline{\text{h}}_{11}\phi^2+\underline{\text{h}}_{22}}\left(\underline{\theta}_1-\theta_1^0\right)-\frac{\phi\underline{\text{h}}_{22}}{\underline{\text{h}}_{11}\phi^2+\underline{\text{h}}_{22}}\left(\underline{\theta}_2-\theta_2^0\right),$$

$$p\lim_{T\to\infty}\left(\bar{\theta}_{2,T}\right)=\theta_2^0-\frac{\phi\underline{\text{h}}_{11}}{\underline{\text{h}}_{11}\phi^2+\underline{\text{h}}_{22}}\left(\underline{\theta}_1-\theta_1^0\right)+\frac{\underline{\text{h}}_{22}}{\underline{\text{h}}_{11}\phi^2+\underline{\text{h}}_{22}}\left(\underline{\theta}_2-\theta_2^0\right)$$

which highlights the role of the prior precisions in the outcomes. In the case where the prior precisions are set to be the same across the parameters and $\underline{\text{h}}_{12}=0$, (often done in practice) we have (10) and (11) above

$$p\lim_{T\to\infty}\left(\bar{\theta}_{1,T}\right)=\theta_1^0+\frac{\phi^2}{1+\phi^2}\left(\underline{\theta}_1-\theta_1^0\right)-\frac{\phi}{1+\phi^2}\left(\underline{\theta}_2-\theta_2^0\right),$$
$$p\lim_{T\to\infty}\left(\bar{\theta}_{2,T}\right)=\theta_2^0-\frac{\phi}{1+\phi^2}\left(\underline{\theta}_1-\theta_1^0\right)+\frac{\phi^2}{1+\phi^2}\left(\underline{\theta}_2-\theta_2^0\right),$$

and the limit of posterior means do not depend on the prior precisions, but do depend on the priors for the coefficients even asymptotically.

A2. Derivation of the posterior mean in the highly collinear case

In the highly collinear case we have

$$\begin{aligned}T^{-1}\mathbf{X}'\mathbf{X}=&\begin{pmatrix} s_T^2 & \phi s_T^2+T^{-1/2}\delta_T\ s_{1v,T}\\ \phi s_T^2+T^{-1/2}\delta_T\ s_{1v,T} & \phi^2 s_T^2+T^{-1}\delta_T^2 s_{vv,T}+2T^{-1/2}\phi\delta_T\ s_{1v,T}\end{pmatrix}\\ =&\ s_T^2\begin{pmatrix}1&\phi\\ \phi&\phi^2\end{pmatrix}+\begin{pmatrix}0 & T^{-1/2}\delta_T\ s_{1v,T}\\ T^{-1/2}\delta_T\ s_{1v,T} & T^{-1}\delta_T^2 s_{vv,T}+2T^{-1/2}\phi\delta_T\ s_{1v,T}\end{pmatrix}.\end{aligned}$$

where

$$s_{1v,T}=T^{-1}\sum_{t=1}^T x_{1t}v_t,\ s_{vv,T}=T^{-1}\sum_{t=1}^T v_t^2.$$

Similarly,

$$T^{-1}\mathbf{X}'\mathbf{y}=\begin{pmatrix}T^{-1}\mathbf{x}_1'\mathbf{y}\\ T^{-1}\mathbf{x}_2'y\end{pmatrix}=\begin{pmatrix}s_T^2\hat{\beta}_T\\ T^{-1}\mathbf{y}'\left(\phi\mathbf{x}_1+\frac{\delta_T}{\sqrt{T}}\mathbf{v}\right)\end{pmatrix}=\begin{pmatrix}s_T^2\hat{\beta}_T\\ s_T^2\phi\hat{\beta}_T+\frac{\delta_T}{\sqrt{T}}T^{-1}\mathbf{y}'\mathbf{v}\end{pmatrix}$$

$$\sigma^{-2}T^{-1}\mathbf{X}'\mathbf{y}=\sigma^{-2}s_T^2\hat{\beta}_T\begin{pmatrix}1\\ \phi\end{pmatrix}+\begin{pmatrix}0\\ \frac{\delta_T}{\sigma^2\sqrt{T}}\left(T^{-1}\mathbf{y}'\mathbf{v}\right)\end{pmatrix},$$

where $s_{yv,T}=T^{-1}\sum_{t=1}^T y_t v_t$. Hence

$$\begin{aligned}\sigma^{-2}T^{-1}\mathbf{X}'\mathbf{X}+T^{-1}\underline{\text{H}}=&\left(s_T^2/\sigma^2\right)\begin{pmatrix}1&\phi\\ \phi&\phi^2\end{pmatrix}+\\ &T^{-1}\begin{pmatrix}\underline{\text{h}}_{11} & \underline{\text{h}}_{12}+\delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right)\\ \underline{\text{h}}_{12}+\delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right) & \underline{\text{h}}_{22}+\delta_T^2\left(s_{vv,T}/\sigma^2\right)+2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\end{pmatrix}\end{aligned}$$

$$\sigma^{-2}T^{-1}\mathbf{X}'\mathbf{y}+T^{-1}\underline{\mathrm{H}}\;\underline{\theta}=\left(s_T^2/\sigma^2\right)\hat{\beta}_T\begin{pmatrix}1\\ \phi\end{pmatrix}+\begin{pmatrix}0\\ \frac{\delta_T}{\sigma^2\sqrt{T}}\left(T^{-1}\sum_{t=1}^T y_t v_t\right)\end{pmatrix}+T^{-1}\begin{pmatrix}b_1\\ b_2\end{pmatrix}$$

$$\sigma^{-2}T^{-1}\mathbf{X}'\mathbf{y}+T^{-1}\underline{\mathrm{H}}\;\underline{\theta}=\left(s_T^2/\sigma^2\right)\hat{\beta}_T\begin{pmatrix}1\\ \phi\end{pmatrix}+T^{-1}\begin{pmatrix}b_1\\ b_2+\frac{\delta_T}{\sigma^2}\left(T^{-1/2}\sum_{t=1}^T y_t v_t\right)\end{pmatrix}$$

$$\begin{aligned}T^{-1/2}\sum_{t=1}^T y_t v_t &= T^{-1/2}\sum_{t=1}^T v_t\left(\theta_1^0 x_{1t}+\theta_2^0\left[\phi x_{1t}+\left(\delta_T/\sqrt{T}\right)v_t\right]+u_t\right)\\ &=\theta_1^0\left(T^{-1/2}\sum_{t=1}^T v_t x_{1t}\right)+\theta_2^0\phi\left(T^{-1/2}\sum_{t=1}^T v_t x_{1t}\right)+\delta_T\theta_2^0\left(T^{-1}\sum_{t=1}^T v_t^2\right)+T^{-1/2}\sum_{t=1}^T v_t u_t\\ &=\beta^0\left(T^{-1/2}\sum_{t=1}^T v_t x_{1t}\right)+\delta_T\theta_2^0\left(T^{-1}\sum_{t=1}^T v_t^2\right)+T^{-1/2}\sum_{t=1}^T v_t u_t.\end{aligned}$$

$$T^{-1/2}\sum_{t=1}^T y_t v_t=\delta_T\theta_2^0 s_{vv,T}+\beta^0 T^{-1/2}\sum_{t=1}^T v_t\left(x_{1t}+u_t\right)$$

$$s_T^2\hat{\beta}_T=T^{-1}\sum_{t=1}^T y_t x_{1t}=T^{-1}\sum_{t=1}^T x_{1t}\left(\theta_1^0 x_{1t}+\theta_2^0\left[\phi x_{1t}+\left(\delta_T/\sqrt{T}\right)v_t\right]+u_t\right),$$

$$\hat{\beta}_T=\beta^0+\frac{\delta_T\theta_2^0}{\sqrt{T}}\left(\frac{s_{1v,T}}{s_T^2}\right)+\frac{s_{1u,T}}{s_T^2}. \tag{32}$$

$$\hat{\beta}_T\rightarrow_p\beta^0=\theta_1^0+\phi\theta_2^0.$$

Consider now the posterior means

$$\bar{\boldsymbol{\theta}}_T=\left[\begin{pmatrix}1&\phi\\ \phi&\phi^2\end{pmatrix}+T^{-1}\begin{pmatrix}a_{11}&a_{12}\\ a_{12}&a_{22}\end{pmatrix}\right]^{-1}\left[\hat{\beta}_T\begin{pmatrix}1\\ \phi\end{pmatrix}+T^{-1}\begin{pmatrix}b_1\\ b_2\end{pmatrix}\right],$$

where the a_{ij} and b_i are now given by

$$\mathbf{A}=\begin{pmatrix}a_{11}&a_{12}\\ a_{12}&a_{22}\end{pmatrix}=\left(\sigma^2/s_T^2\right)\begin{pmatrix}\underline{\mathrm{h}}_{11} & \underline{\mathrm{h}}_{12}+\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\\ \underline{\mathrm{h}}_{12}+\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right) & \underline{\mathrm{h}}_{22}+\delta_T^2\left(s_{vv,T}/\sigma^2\right)+2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\end{pmatrix},$$

$$\mathbf{b}=\begin{pmatrix}b_1\\ b_2\end{pmatrix}=\left(\sigma^2/s_T^2\right)\underline{\mathrm{H}}\;\underline{\theta}=\left(\sigma^2/s_T^2\right)\begin{pmatrix}\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\\ \underline{\mathrm{h}}_{12}\underline{\theta}_1+\underline{\mathrm{h}}_{22}\underline{\theta}_2+\frac{\delta_T}{\sigma^2}\left(T^{-1/2}\sum_{t=1}^T y_t v_t\right)\end{pmatrix}$$

$$\bar{\boldsymbol{\theta}}_T=\begin{pmatrix}\frac{1}{a_{11}\phi^2-2\phi a_{12}+a_{22}+T^{-1}\left[a_{11}a_{22}-a_{12}^2\right]}\left(\hat{\beta}_T\left(a_{22}-\phi a_{12}\right)+\phi\left(\phi b_1-b_2\right)+T^{-1}\left(b_1a_{22}-b_2a_{12}\right)\right)\\ \frac{1}{a_{11}\phi^2-2\phi a_{12}+a_{22}+T^{-1}\left[a_{11}a_{22}-a_{12}^2\right]}\left(b_2-\phi b_1-\hat{\beta}_T\left(a_{12}-\phi a_{11}\right)+T^{-1}\left(b_2a_{11}-b_1a_{12}\right)\right)\end{pmatrix},$$

To evaluate this first consider the denominator of $\bar{\theta}_{1,T}$, where both numerator and denominator are multiplied by $\left(\sigma^2/s_T^2\right)^{-1}$

$$\left(\sigma^2/s_T^2\right)^{-1}\left[\hat{\beta}_T\left(a_{22}-\phi a_{12}\right)+\phi\left(\phi b_1-b_2\right)+T^{-1}\left(b_1a_{22}-b_2a_{12}\right)\right]$$

$$=\left[\beta^0+\frac{\delta_T\theta_2^0}{\sqrt{T}}\left(\frac{s_{1v,T}}{s_T^2}\right)+\frac{s_{1u,T}}{s_T^2}\right]\left[\underline{\mathrm{h}}_{22}+\delta_T^2\left(s_{vv,T}/\sigma^2\right)+2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)-\phi\underline{\mathrm{h}}_{12}-\phi\delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right)\right]$$

$$+\phi^2\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\right)-\phi\left[\underline{\mathrm{h}}_{12}\underline{\theta}_1+\underline{\mathrm{h}}_{22}\underline{\theta}_2+\frac{\delta_T}{\sigma^2}\left[\delta_T\theta_2^0s_{vv,T}+\beta^0T^{-1/2}\sum_{t=1}^{T}v_t\left(x_{1t}+u_t\right)\right]\right]$$

$$+\left(\sigma^2/s_T^2\right)T^{-1}\left[\begin{array}{c}\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\right)\left[\underline{\mathrm{h}}_{22}+\delta_T^2\left(s_{vv,T}/\sigma^2\right)+2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\right]\\-\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1+\underline{\mathrm{h}}_{22}\underline{\theta}_2+\frac{\delta_T}{\sigma^2}\left(T^{-1/2}\sum_{t=1}^{T}y_tv_t\right)\right)\left(\underline{\mathrm{h}}_{12}+\delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right)\right)\end{array}\right]$$

$$=\left[\beta^0+\frac{\delta_T\theta_2^0}{\sqrt{T}}\left(\frac{s_{1v,T}}{s_T^2}\right)+\frac{s_{1u,T}}{s_T^2}\right]\left[\underline{\mathrm{h}}_{22}-\phi\underline{\mathrm{h}}_{12}+\delta_T^2\left(s_{vv,T}/\sigma^2\right)+\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\right]$$

$$+\phi^2\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\right)-\phi\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1+\underline{\mathrm{h}}_{22}\underline{\theta}_2\right)-\frac{\phi\delta_T}{\sigma^2}\left[\delta_T\theta_2^0s_{vv,T}+\beta^0T^{-1/2}\sum_{t=1}^{T}v_t\left(x_{1t}+u_t\right)\right]$$

$$+\left(\sigma^2/s_T^2\right)T^{-1}\left[\begin{array}{c}\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\right)\left[\underline{\mathrm{h}}_{22}+\delta_T^2\left(s_{vv,T}/\sigma^2\right)\right]+2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\right)\\-\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1+\underline{\mathrm{h}}_{22}\underline{\theta}_2+\frac{\delta_T}{\sigma^2}\left(T^{-1/2}\sum_{t=1}^{T}y_tv_t\right)\right)\left(\underline{\mathrm{h}}_{12}+\delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right)\right)\end{array}\right]$$

$$=\left[\beta^0+\frac{\delta_T\theta_2^0}{\sqrt{T}}\left[\frac{s_{1v,T}}{s_T^2}\right]+\frac{s_{1u,T}}{s_T^2}\right]\left[\underline{\mathrm{h}}_{22}-\phi\underline{\mathrm{h}}_{12}+\delta_T^2\left(s_{vv,T}/\sigma^2\right)\right]$$

$$+\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\frac{s_{1u,T}}{s_T^2}+\phi\delta_T^2\theta_2^0\left(s_{1v,T}/\sigma^2\right)\left[\frac{s_{1v,T}}{s_T^2}\right]$$

$$+\phi^2\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\right)-\phi\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1+\underline{\mathrm{h}}_{22}\underline{\theta}_2\right)$$

$$-\frac{\phi\delta_T}{\sigma^2}\delta\theta_2^0s_{vv,T}-\frac{\phi\delta_T}{\sigma^2}\beta^0T^{-1/2}\sum_{t=1}^{T}v_t\left(x_{1t}+u_t\right)+\beta^0\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)$$

$$+\left(\sigma^2/s_T^2\right)T^{-1}\left[\begin{array}{c}\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\right)\left[\underline{\mathrm{h}}_{22}+\delta_T^2\left(s_{vv,T}/\sigma^2\right)\right]+2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\right)\\-\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1+\underline{\mathrm{h}}_{22}\underline{\theta}_2+\frac{\delta_T}{\sigma^2}\left(T^{-1/2}\sum_{t=1}^{T}y_tv_t\right)\right)\left(\underline{\mathrm{h}}_{12}+\delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right)\right)\end{array}\right]$$

$$=\left[\beta^0+\frac{\delta_T\theta_2^0}{\sqrt{T}}\left[\frac{s_{1v,T}}{s_T^2}\right]+\frac{s_{1u,T}}{s_T^2}\right]\left[\underline{\mathrm{h}}_{22}-\phi\underline{\mathrm{h}}_{12}+\delta_T^2\left(s_{vv,T}/\sigma^2\right)\right]+\phi\delta_T\left(\frac{T^{1/2}s_{1v,T}s_{1u,T}}{\sigma^2s_T^2}\right)+\phi\delta_T^2\theta_2^0\left(\frac{s_{1v,T}^2}{\sigma^2s_T^2}\right)$$

$$+\phi^2\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\right)-\phi\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1+\underline{\mathrm{h}}_{22}\underline{\theta}_2\right)-\frac{\phi\delta^2}{\sigma^2}\theta_2^0s_{vv,T}-\frac{\beta\phi\delta}{\sigma^2}\left(T^{-1/2}\sum_{t=1}^{T}v_tu_t\right)$$

$$+\left(\sigma^2/s_T^2\right)T^{-1}\left[\begin{array}{c}\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\right)\left[\underline{\mathrm{h}}_{22}+\delta^2\left(s_{vv,T}/\sigma^2\right)\right]+2\phi\delta\left(T^{1/2}s_{1v,T}/\sigma^2\right)\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1+\underline{\mathrm{h}}_{12}\underline{\theta}_2\right)\\-\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1+\underline{\mathrm{h}}_{22}\underline{\theta}_2+\frac{\delta}{\sigma^2}\left(T^{-1/2}\sum_{t=1}^{T}y_tv_t\right)\right)\left(\underline{\mathrm{h}}_{12}+\delta\ \left(T^{1/2}s_{1v,T}/\sigma^2\right)\right)\end{array}\right]$$

$$= \left[\frac{\delta_T \theta_2^0}{\sqrt{T}}\left(\frac{s_{1v,T}}{s_T^2}\right) + \frac{s_{1u,T}}{s_T^2}\right]\left[\underline{\mathrm{h}}_{22} - \phi\underline{\mathrm{h}}_{12} + \delta_T^2\left(s_{vv,T}/\sigma^2\right)\right] + \beta\left(\underline{\mathrm{h}}_{22} - \phi\underline{\mathrm{h}}_{12}\right)$$

$$+ \phi\delta\left(\frac{T^{1/2}s_{1v,T}s_{1u,T}}{\sigma^2 s_T^2}\right) + \phi\delta_T^2\theta_2^0\left(\frac{s_{1v,T}^2}{\sigma^2 s_T^2}\right) - \frac{\phi\delta_T^2}{\sigma^2}\theta_2^0 s_{vv,T} + \delta_T^2\beta\left(s_{vv,T}/\sigma^2\right)$$

$$\phi^2\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1 + \underline{\mathrm{h}}_{12}\underline{\theta}_2\right) - \phi\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1 + \underline{\mathrm{h}}_{22}\underline{\theta}_2\right) - \frac{\beta\phi\delta_T}{\sigma^2}\left(T^{-1/2}\sum_{t=1}^T v_t u_t\right)$$

$$+ \left(\sigma^2/s_T^2\right)T^{-1}\left[\begin{array}{c}\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1 + \underline{\mathrm{h}}_{12}\underline{\theta}_2\right)\left[\underline{\mathrm{h}}_{22} + \delta_T^2\left(s_{vv,T}/\sigma^2\right)\right] + 2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1 + \underline{\mathrm{h}}_{12}\underline{\theta}_2\right) \\ -\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1 + \underline{\mathrm{h}}_{22}\underline{\theta}_2 + \frac{\delta_T}{\sigma^2}\left(T^{-1/2}\sum_{t=1}^T y_t v_t\right)\right)\left(\underline{\mathrm{h}}_{12} + \delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right)\right)\end{array}\right]$$

$$= \left[\frac{\delta_T \theta_2^0}{\sqrt{T}}\left[\frac{s_{1v,T}}{s_T^2}\right] + \frac{s_{1u,T}}{s_T^2}\right]\left[\underline{\mathrm{h}}_{22} - \phi\underline{\mathrm{h}}_{12} + \delta_T^2\left(s_{vv,T}/\sigma^2\right)\right]$$

$$+ \phi\delta_T\left(\frac{T^{1/2}s_{1v,T}s_{1u,T}}{\sigma^2 s_T^2}\right) + \phi\delta_T^2\theta_2^0\left(\frac{s_{1v,T}^2}{\sigma^2 s_T^2}\right) + \delta_T^2\theta_1\left(s_{vv,T}/\sigma^2\right)$$

$$\phi^2\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1 + \underline{\mathrm{h}}_{12}\underline{\theta}_2\right) - \phi\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1 + \underline{\mathrm{h}}_{22}\underline{\theta}_2\right) + \beta\left(\underline{\mathrm{h}}_{22} - \phi\underline{\mathrm{h}}_{12}\right) - \frac{\beta\phi\delta_T}{\sigma^2}\left(T^{-1/2}\sum_{t=1}^T v_t u_t\right)$$

$$+ \left(\sigma^2/s_T^2\right)T^{-1}\left[\begin{array}{c}\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1 + \underline{\mathrm{h}}_{12}\underline{\theta}_2\right)\left[\underline{\mathrm{h}}_{22} + \delta_T^2\left(s_{vv,T}/\sigma^2\right)\right] + 2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1 + \underline{\mathrm{h}}_{12}\underline{\theta}_2\right) \\ -\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1 + \underline{\mathrm{h}}_{22}\underline{\theta}_2 + \frac{\delta_T}{\sigma^2}\left(T^{-1/2}\sum_{t=1}^T y_t v_t\right)\right)\left(\underline{\mathrm{h}}_{12} + \delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right)\right)\end{array}\right]$$

When δ_T is bounded in T we have

$$\begin{aligned}\bar{\theta}_{1,T} = &-\frac{\beta\phi\delta_T}{\sigma^2\left[\underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} + \frac{\delta_T^2 s_{vv,T}}{\sigma^2}\right]}\left(T^{-1/2}\sum_{t=1}^T v_t u_t\right) + \\ &\frac{\phi^2\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1 + \underline{\mathrm{h}}_{12}\underline{\theta}_2\right) - \phi\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1 + \underline{\mathrm{h}}_{22}\underline{\theta}_2\right) + \beta\left(\underline{\mathrm{h}}_{22} - \phi\underline{\mathrm{h}}_{12}\right)}{\underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} + \frac{\delta_T^2 s_{vv,T}}{\sigma^2}} \\ &+ \frac{\delta_T^2\theta_1\left(s_{vv,T}/\sigma^2\right)}{\underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} + \frac{\delta_T^2 s_{vv,T}}{\sigma^2}} + O_p\left(T^{-1/2}\right)\end{aligned}$$

Similarly, for the denominator we note that

$$\begin{aligned}
&\left(\sigma^2/s_T^2\right)^{-1}\left\{a_{11}\phi^2 - 2\phi a_{12} + a_{22} + T^{-1}\left[a_{11}a_{22} - a_{12}^2\right]\right\} \\
&= \underline{\mathrm{h}}_{11}\phi^2 - 2\phi\left[\underline{\mathrm{h}}_{12} + \delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right)\right] + \underline{\mathrm{h}}_{22} + \delta_T^2\left(s_{vv,T}/\sigma^2\right) + 2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right) \\
&+ T^{-1}\left(\sigma^2/s_T^2\right)\left[\underline{\mathrm{h}}_{11}\left[\underline{\mathrm{h}}_{22} + \delta_T^2\left(s_{vv,T}/\sigma^2\right) + 2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\right] - \left[\underline{\mathrm{h}}_{12} + \delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right)\right]^2\right] \\
&= \underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} - 2\phi\delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right) + \underline{\mathrm{h}}_{22} + \delta_T^2\left(s_{vv,T}/\sigma^2\right) + 2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right) \\
&+ T^{-1}\left(\sigma^2/s_T^2\right)\left[\underline{\mathrm{h}}_{11}\left[\underline{\mathrm{h}}_{22} + \delta_T^2\left(s_{vv,T}/\sigma^2\right) + 2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\right] - \left[\underline{\mathrm{h}}_{12} + \delta_T\ \left(T^{1/2}s_{1v,T}/\sigma^2\right)\right]^2\right] \\
&= \underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} + \delta_T^2\left(s_{vv,T}/\sigma^2\right) \\
&+ T^{-1}\left(\sigma^2/s_T^2\right)\left[\begin{array}{c} \underline{\mathrm{h}}_{11}\underline{\mathrm{h}}_{22} + \delta_T^2\left(s_{vv,T}/\sigma^2\right)\underline{\mathrm{h}}_{11} + 2\phi\delta_T\left(T^{1/2}s_{1v,T}/\sigma^2\right)\underline{\mathrm{h}}_{11} \\ -\underline{\mathrm{h}}_{12}^2 - \delta_T^2 T\left(s_{1v,T}/\sigma^2\right)^2 - 2\delta_T\underline{\mathrm{h}}_{12}\left(T^{1/2}s_{1v,T}/\sigma^2\right) \end{array}\right] \\
&= \underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} + \delta_T^2\left(s_{vv,T}/\sigma^2\right) - \delta_T^2\left(s_{1v,T}/\sigma^2\right)^2\left(\sigma^2/s_T^2\right) \\
&+ 2T^{-1/2}\phi\delta_T\left(s_{1v,T}/\sigma^2\right)\underline{\mathrm{h}}_{11}\left(\sigma^2/s_T^2\right) - 2\delta_T\underline{\mathrm{h}}_{12}\left(s_{1v,T}/\sigma^2\right)\left(\sigma^2/s_T^2\right) \\
&+ T^{-1}\left(\sigma^2/s_T^2\right)\left[\underline{\mathrm{h}}_{11}\underline{\mathrm{h}}_{22} + \delta_T^2\left(s_{vv,T}/\sigma^2\right)\underline{\mathrm{h}}_{11} - \underline{\mathrm{h}}_{12}^2\right]
\end{aligned}$$

Or

$$\begin{aligned}
&\left(\sigma^2/s_T^2\right)^{-1}\left\{a_{11}\phi^2 - 2\phi a_{12} + a_{22} + T^{-1}\left[a_{11}a_{22} - a_{12}^2\right]\right\} \\
&= \underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} + \delta_T^2\left[\frac{s_{vv,T}s_T^2 - s_{1v,T}^2}{\sigma^2 s_T^2}\right] \\
&+ 2T^{-1/2}\delta_T\left(s_{1v,T}/s_T^2\right)\left[\phi\underline{\mathrm{h}}_{11} - \underline{\mathrm{h}}_{12}\right] \\
&+ T^{-1}\left(\sigma^2/s_T^2\right)\left[\underline{\mathrm{h}}_{11}\underline{\mathrm{h}}_{22} + \delta_T^2\left(s_{vv,T}/\sigma^2\right)\underline{\mathrm{h}}_{11} - \underline{\mathrm{h}}_{12}^2\right]
\end{aligned}$$

In the case where δ_T is bounded in T we obtain

$$\begin{aligned}
&\left(\sigma^2/s_T^2\right)^{-1}\left\{a_{11}\phi^2 - 2\phi a_{12} + a_{22} + T^{-1}\left[a_{11}a_{22} - a_{12}^2\right]\right\} \\
&= \underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} + \frac{\delta_T^2 s_{vv,T}}{\sigma^2} + O_p(T^{-1})
\end{aligned}$$

But $s_{1v,T} = O_p(T^{-1/2})$, and $s_{vv,T} = \sigma_v^2 + O_p(T^{-1})$

$$\begin{aligned}
\bar{\theta}_{1,T} = &\frac{\phi^2\left(\underline{\mathrm{h}}_{11}\underline{\theta}_1 + \underline{\mathrm{h}}_{12}\underline{\theta}_2\right) - \phi\left(\underline{\mathrm{h}}_{12}\underline{\theta}_1 + \underline{\mathrm{h}}_{22}\underline{\theta}_2\right) + \beta^0\left(\underline{\mathrm{h}}_{22} - \phi\underline{\mathrm{h}}_{12}\right)}{\underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} + \left(\frac{\delta_T^2\sigma_v^2}{\sigma^2}\right)} \\
&+ \frac{\theta_1^0\left(\frac{\delta_T^2\sigma_v^2}{\sigma^2}\right)}{\underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} + \left(\frac{\delta_T^2\sigma_v^2}{\sigma^2}\right)} \\
&+ \frac{-\beta^0\phi\sigma_v\sigma\delta_T}{\sigma^2\left[\underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} + \left(\frac{\delta_T^2\sigma_v^2}{\sigma^2}\right)\right]}\left(T^{-1/2}\sum_{t=1}^{T}\frac{v_t u_t}{\sigma_v\sigma}\right) + O_p\left(T^{-1/2}\right).
\end{aligned}$$

The above results can be simplified further by setting $\lambda_T^2 = \delta_T^2\sigma_v^2/\sigma^2$, and noting that $\beta^0 =$

$\theta_1^0 + \phi\theta_2^0$, $\underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22} = \psi'\underline{\mathbf{H}}\psi$, where $\boldsymbol{\psi} = (\phi, -1)'$. Specifically,

$$\bar{\theta}_{1,T} = \theta_1^0 + \frac{\phi\left(\underline{\mathrm{h}}_{11}\phi - \underline{\mathrm{h}}_{12}\right)\left(\underline{\theta}_1 - \theta_1^0\right)}{\lambda_T^2 + \boldsymbol{\psi}'\underline{\mathbf{H}}\boldsymbol{\psi}} - \frac{\phi\left(\underline{\mathrm{h}}_{22} - \phi\underline{\mathrm{h}}_{12}\right)\left(\underline{\theta}_2 - \theta_2^0\right)}{\lambda_T^2 + \boldsymbol{\psi}'\underline{\mathbf{H}}\boldsymbol{\psi}} - \left(\frac{\beta^0\phi\lambda_T}{\lambda_T^2 + \boldsymbol{\psi}'\underline{\mathbf{H}}\boldsymbol{\psi}}\right)\left(T^{-1/2}\sum_{t=1}^{T}\frac{v_t u_t}{\sigma_v\sigma}\right) + O_p\left(T^{-1/2}\right). \tag{33}$$

Thus as $T \to \infty$, in the highly collinear case where λ_T is bounded in T, the posterior mean, $\bar{\theta}_{1,T}$, converges in distribution to a normally distributed random variable given in subsection 3.1.

A3. Derivation of posterior precision in the highly collinear case

Starting with (17) we note that $\bar{\mathbf{V}}^{-1}$ can be written as

$$\bar{\mathbf{V}}^{-1} = \tilde{s}_T^2\begin{pmatrix} T & T\phi \\ T\phi & T\phi^2 \end{pmatrix} + \begin{pmatrix} \underline{\mathrm{h}}_{11} & \underline{\mathrm{h}}_{12} + \chi_T z_T \\ \underline{\mathrm{h}}_{12} + \chi_T z_T & \underline{\mathrm{h}}_{22} + \lambda_T^2 + 2\chi_T\phi z_T \end{pmatrix},$$

where

$$\tilde{s}_T^2 = s_T^2/\sigma^2,\ \lambda_T^2 = \delta_T^2\left(s_{vv,T}/\sigma^2\right),\ \chi_T = \frac{\delta_T\sigma_v\sigma_{x_1}}{\sigma^2},$$

$$z_T = \frac{T^{1/2}s_{1v,T}}{\sigma_{x_1}\sigma_v} = T^{-1/2}\sum_{t=1}^{T}\frac{x_{1t}v_t}{\sigma_{x_1}\sigma_v}.$$

Hence

$$\bar{\mathbf{V}}^{-1} = \begin{pmatrix} \underline{\mathrm{h}}_{11} + T\tilde{s}_T^2 & T\phi\tilde{s}_T^2 + \underline{\mathrm{h}}_{12} + \chi_T z_T \\ T\phi\tilde{s}_T^2 + \underline{\mathrm{h}}_{12} + \chi_T z_T & T\phi^2\tilde{s}_T^2 + \underline{\mathrm{h}}_{22} + \lambda_T^2 + 2\chi_T\phi z_T \end{pmatrix},$$

and the posterior precision of θ_1 is given by the inverse of the first element of $\bar{\mathbf{V}}$, which is given by

$$\begin{aligned}
\bar{h}_{11,T} &= \underline{\mathrm{h}}_{11} + T\tilde{s}_T^2 - \frac{\left(T\phi\tilde{s}_T^2 + \underline{\mathrm{h}}_{12} + \chi_T z_T\right)^2}{T\phi^2\tilde{s}_T^2 + \underline{\mathrm{h}}_{22} + \lambda_T^2 + 2\chi_T\phi z_T} \\
&= \frac{\left(\underline{\mathrm{h}}_{11} + T\tilde{s}_T^2\right)\left(T\phi^2\tilde{s}_T^2 + \underline{\mathrm{h}}_{22} + \lambda_T^2 + 2\chi_T\phi z_T\right) - \left(T\phi\tilde{s}_T^2 + \underline{\mathrm{h}}_{12} + \chi_T z_T\right)^2}{T\phi^2\tilde{s}_T^2 + \underline{\mathrm{h}}_{22} + \lambda_T^2 + 2\chi_T\phi z_T} \\
&= \frac{\begin{array}{c}\underline{\mathrm{h}}_{11}\left(T\phi^2\tilde{s}_T^2 + \underline{\mathrm{h}}_{22} + \lambda_T^2 + 2\chi_T\phi z_T\right) + T\tilde{s}_T^2\left(\underline{\mathrm{h}}_{22} + \lambda_T^2 + 2\chi_T\phi z_T\right) \\ -\underline{\mathrm{h}}_{12}^2 - \chi_T^2 z_T^2 - 2\underline{\mathrm{h}}_{12}\chi_T z_T - 2T\phi\tilde{s}_T^2\left(\underline{\mathrm{h}}_{12} + \chi_T z_T\right)\end{array}}{T\phi^2\tilde{s}_T^2 + \underline{\mathrm{h}}_{22} + \lambda_T^2 + 2\chi_T\phi z_T} \\
&= \frac{\begin{array}{c}T\underline{\mathrm{h}}_{11}\phi^2\tilde{s}_T^2 + \underline{\mathrm{h}}_{11}\lambda_T^2 + 2\underline{\mathrm{h}}_{11}\chi_T\phi z_T + T\tilde{s}_T^2\underline{\mathrm{h}}_{22} + T\tilde{s}_T^2\lambda_T^2 \\ \underline{\mathrm{h}}_{11}\underline{\mathrm{h}}_{22} - \underline{\mathrm{h}}_{12}^2 - \chi_T^2 z_T^2 - 2\underline{\mathrm{h}}_{12}\chi_T z_T - 2T\phi\tilde{s}_T^2\underline{\mathrm{h}}_{12}\end{array}}{T\phi^2\tilde{s}_T^2 + \underline{\mathrm{h}}_{22} + \lambda_T^2 + 2\chi_T\phi z_T}.
\end{aligned}$$

Or

$$\bar{h}_{11,T} = \frac{\begin{array}{c}T\tilde{s}_T^2\left(\lambda_T^2 + \underline{\mathrm{h}}_{11}\phi^2 - 2\phi\underline{\mathrm{h}}_{12} + \underline{\mathrm{h}}_{22}\right) - \chi_T^2 z_T^2 + 2\chi_T\left(\underline{\mathrm{h}}_{11}\phi - \underline{\mathrm{h}}_{12}\right)z_T + \\ +\underline{\mathrm{h}}_{11}\lambda_T^2 + \underline{\mathrm{h}}_{11}\underline{\mathrm{h}}_{22} - \underline{\mathrm{h}}_{12}^2\end{array}}{T\phi^2\tilde{s}_T^2 + 2\chi_T\phi z_T + \underline{\mathrm{h}}_{22} + \lambda_T^2}$$

from which the expression in the text, (18), for the posterior precision of θ_1 follows.

References

Basturk, N., L. Hoogerheide and H. van Dijk (2017) Bayesian analysis of boundary and near-boundary evidence in econometric models with reduced rank, *Bayesian Analysis*, 12(3), 879-917..

Baumeister, C. and J.D. Hamilton (2015) Sign restrictions, structural vector autoregressions and useful prior information, *Econometrica*, 85(5) 1963-1999.

Cheng, T., J. Gao and P. C. B. Phillips (2017) Bayesian estimation based on summary statistics: double asymptotics and practice. Monash Working paper 4/17.

Frisch, R. (1934) *Statistical Confluence Analysis by means of complete regression systems*, University Institute of Economics, Olso.

Klein, L.R. (1962) *An Introduction to Econometrics*, Prentice Hall.

Koop G., M.H. Pesaran and R.P. Smith (2013), On Identification of Bayesian DSGE Models, *Journal of Business and Economic Statistics*, 31, pp. 300-314.

Leamer, E. (1978) *Specification Searches: ad hoc inference with non-experimental data*, Wiley.

Pesaran, M.H. (2015) *Time Series and Panel Data Econometrics*, Oxford University Press.

Phillips, P. C. B., (2016), Inference in Near-Singular Regression, in Gloria GonzÁlez-Rivera , R. Carter Hill , Tae-Hwy Lee (ed.) *Essays in Honor of Aman Ullah* (Advances in Econometrics, Volume 36) Emerald Group Publishing Limited, pp.461 - 486.

Poirier, D.J. (1998) Revising beliefs in nonidentified models, *Econometric Theory* 14, 483-509.

Sanderson, E. and F. Windmeijer (2016) A weak instrument F-test in linear IV models with multiple endogenous variables, *Journal of Econometrics*, 190, 212-221.

Spanos, A. and A. McGuirk (2002) The problem of near-multicollinearity revisited: erratic vs systematic volatility, *Journal of Econometrics* 108, 365-393.

Staiger, D. and J.H. Stock (1997) Instrumental variable regression with weak instruments, *Econometrica*, 65, 557-586.

Stambaugh, R.F., (1999) Predictive Regressions, *Journal of Financial Economics*, 54(3), 375-421.

Stock, J.H., J. H. Wright, and M. Yogo (2002) A survey of weak instruments and weak identification in generalized method of moments, *Journal of Business and Economic Statistics*, 20, 518-529

Tobin, J. (1950) A statistical demand function for food in the USA, *Journal of the Royal Statistical Society*, series A, 113-141.