Noname manuscript No.
(will be inserted by the editor)

Improvement in Finite-Sample Properties of GMM-Based Wald Tests

Qihui Chen · Yu Ren

Received: date / Accepted: date

Abstract GMM-based Wald tests tend to overreject when used for small samples, mainly due to inaccurate estimation of the weighting matrix. This article proposes applying the shrinkage method to address this problem. Using a possibly-misspecified factor model, the shrinkage method can provide a good estimator for the weighting matrix, and hence improve the finite-sample performance of the GMM-based Wald tests.

Keywords Generalized method of moments · Wald tests · Finite-sample properties · Covariance matrix · Shrinkage method

1 Introduction

The generalized method of moments (GMM) procedures outlined by Hansen (1982) are widely-used in economics. The estimation is conducted by minimizing a criterion function which combines all the available moment conditions in a quadratic form with certain weights. As for the hypothesis testing, the GMM-based Wald statistic is the distance measure for the degree to which the

We thank the editor, the associate editor and the referee for their comments. We also thank Ying Shang for her help. Ren's research was supported by Xiamen University, the Natural Science Foundation of Fujian Province of China (Grant no. 2011J01384) and the Natural Science Foundation of China (Grant no. 71131008 and no. 70971113).

Y. Ren (Corresponding author)
WISE, Economics Building
MOE Key Lab of Econometrics and Fujian Key Lab of Statistics
Xiamen University, Fujian, China 361005.
Tel.: +86-592-218-6025
Fax: +86-592-218-7708
E-mail: renxmu@gmail.com

Q. Chen
Department of Economics, University of California, San Diego, CA 92093, USA

unrestricted estimator fails to satisfy the restrictions. Most inferences of Wald tests are based on their asymptotic properties. However, many researchers have mentioned that GMM-based Wald tests tend to overreject moment restrictions in finite-samples (for example, Rothenberg (1984), Christiano and den Haan (1996), Kan and Zhang (1999) and Bekaert and Hodrick (2001)). As for potential solutions to remedy the Wald tests' limitation, Hansen et al. (1996) use an iterative weighting matrix; Andersen and Sørenson (1996) emphasize controlling the number of moments; Burnside and Eichenbaum (1996) suggest imposing the null hypotheses into the estimation of the weighting matrix; Hall and Horowitz (1996) propose using a bootstrap approach to derive the inferences. Although their solutions are different, they all find that the poor small-sample performance of the GMM-based Wald tests is due to the inaccurate estimation of the weighting matrix. The weighting matrix of the Wald test involves an asymptotic covariance matrix for the moments. When the sample size is small, relative to the number of moments, the sample covariance matrix is not accurate.

This paper proposes using the shrinkage method to estimate this covariance matrix in small samples. If the moments of interest can be related with a factor model, a good estimator of the covariance matrix can be obtained by following the shrinkage methods in Ledoit and Wolf (2003) and Ren and Shimotsu (2009). The basic idea behind shrinkage estimation is to take an optimally weighted average of the sample covariance matrix and the covariance matrix implied by a possibly-misspecified factor model. On the one hand, the factor model provides a covariance matrix estimate that is biased but has only a small estimation error due to the small number of parameters estimated. On the other hand, the sample covariance matrix provides another estimate which has a small bias, but a large estimation error. Thus, the shrinkage estimator balances the trade-off between the estimation error and bias by taking a weighted average of these two estimates.

The shrinkage method is valid only when the factor model, even misspecified, exists. This condition is quite mild in practice. Many variables in macroeconomics or finance have this property. Compared with other estimators mentioned in the literature, the shrinkage estimator has three advantages. First, it is easy to compute. It is a one-step estimator. Second, unlike the estimator in Burnside and Eichenbaum (1996), it is robust to modification of the null hypothesis. Last, it is more reliable because it adopts more information from the factor model. So if we use this estimator as the weighting matrix for Wald tests, their performance should be superior to the alternatives.

Following Burnside and Eichenbaum (1996), we conduct several simulations to support our conjecture. We compare the small-sample performance of Wald tests by adopting different estimations of the weighting matrix. The shrinkage method outperforms all others. In addition, as a simple illustration of the empirical importance of our method, we revisit the mean-variance efficiency of portfolio returns. We collect the Fama-French 25 portfolios and follow Markowitz (1952) to derive one mean-variance efficient portfolio. Furthermore, we generate some portfolios which have the same mean but a different variance from the efficient portfolio. We test the null hypothesis that the volatilities of the generated portfolios are the same. If we use the conventional Wald tests, we can reject the null at the 5 percent significance level. However, with our approach, we draw the opposite conclusion.

The remainder of this paper is organized as follows. Section 2 briefly reviews the shrinkage method. Sections 3 and 4 show our simulation results and our empirical analysis, with a conclusion in Section 5.

2 Shrinkage Estimation

Suppose $\mathbf{X}_t = [X_{1t}\ X_{2t}\ \dots\ X_{Jt}]'$, where $t = 1, 2, \cdots, T$, is a random vector with zero mean. We define $\mathbf{X}_t^2 = [X_{1t}^2\ X_{2t}^2\ \dots\ X_{Jt}^2]'$. Following Burnside and Eichenbaum (1996), we are interested in the covariance matrix of $\mathbf{X}_t^2$, which is denoted by $\mathbf{S}_0$. In our context, the sample covariance matrix, denoted by $\mathbf{S}_T^{(1)}$, is an unbiased estimator of $\mathbf{S}_0$ but has a large variance. Suppose $\mathbf{X}_t^2$ can be expressed by a factor model as

$$\mathbf{X}_t^2 = \boldsymbol{\eta} + \boldsymbol{\beta}\mathbf{V_t} + \boldsymbol{\epsilon_t}, \tag{1}$$

where $\boldsymbol{\eta} = [\eta_1\ \eta_2\ \cdots\ \eta_J]'$, $\boldsymbol{\epsilon_t} = [\epsilon_{1t}\ \epsilon_{2t}\ \dots\ \epsilon_{Jt}]'$, $\mathbf{V_t}$ is a $K \times 1$ factor, $\boldsymbol{\beta} = [\boldsymbol{\beta_1}\ \boldsymbol{\beta_2}\ \cdots\ \boldsymbol{\beta_J}]'$ and $\boldsymbol{\beta_i}$ is a $K \times 1$ vector. Here, no correlation between ϵ_{it} and ϵ_{jt} for any $i \neq j$ is assumed.

This factor model implies that the covariance matrix of $\mathbf{X}_t^2$, $\boldsymbol{\Phi_0}$, is given by

$$\boldsymbol{\Phi_0} = \boldsymbol{\beta}\mathrm{Var}(\mathbf{V_t})\boldsymbol{\beta}' + \boldsymbol{\Delta}, \tag{2}$$

where $\boldsymbol{\Delta} = \mathrm{diag}\{\mathrm{Var}(\epsilon_{1t}), \mathrm{Var}(\epsilon_{2t}), \cdots, \mathrm{Var}(\epsilon_{Jt})\}$. The matrix $\boldsymbol{\Phi_0}$ can be estimated by estimating its components using the least squares method. Let $\mathbf{b}$ and $\mathbf{D}$ be the estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\Delta}$, respectively. Thus, the estimator of $\boldsymbol{\Phi_0}$ is

$$\boldsymbol{\Phi_T} = \mathbf{b}\widehat{\mathrm{Var}(\mathbf{V_t})}\mathbf{b}' + \mathbf{D}, \tag{3}$$

where $\widehat{\mathrm{Var}(\mathbf{V_t})}$ is the sample covariance matrix of $\mathbf{V_t}$.

The shrinkage estimator takes an optimally weighted average of the sample covariance matrix $\mathbf{S}_T^{(1)}$ and $\boldsymbol{\Phi_T}$ as

$$\mathbf{S}_T^{(s)} = \hat{a}_T\mathbf{\Phi}_T + (1 - \hat{a}_T)\mathbf{S}_T^{(1)}. \tag{4}$$

Here, $\hat{a}_T$ is the estimator of the optimal weight a_T[1], which minimizes the expected distance between $\mathbf{S}_T^{(s)}$ and the true covariance matrix $\mathbf{S}_0$. If we use the Frobenius norm [2] to denote this distance, a_T can be expressed as

$$a_T = \arg\min_{a_T} E||a_T\mathbf{\Phi}_T + (1 - a_T)\mathbf{S}_T^{(1)} - \mathbf{S}_0||. \tag{5}$$

[1] For more details about computing $\hat{a}_T$, please refer to Eq. (10) in Ren and Shimotsu (2009).

[2] For an $N \times N$ matrix $\mathbf{Z}$, the Frobenius norm is $||\mathbf{Z}|| = \sum_{i=1}^N \sum_{j=1}^N z_{ij}^2$.

We assume that $\mathbf{X}_t^2$ and $\mathbf{V_t}$ are independently and identically distributed (i.i.d.) over time[3] and that $\mathbf{X}_t^2$ and $\mathbf{V_t}$ have a finite fourth moment. Based on the corollary in Ren and Shimotsu (2009), $\mathbf{S}_T^{(s)}$ is a consistent estimator of $\mathbf{S}_0$.

3 Simulation

Burnside and Eichenbaum (1996) study a very simple case, in which the time series X_{it} is mutually independent. We are going to relax this strong assumption, and suppose that we have time series data generated by the data-generating process (DGP)

$$X_{it} = \sqrt{r} v_t + \sqrt{1-r} e_{it} \text{ for all } i \text{ and } t, \tag{6}$$

where the common factor v_t is i.i.d. $N(0,1)$ for all t, and e_{it} is mutually and serially independent $N(0,1)$. Thus, X_{it} still follows $N(0,1)$ and the correlation between X_{it} and X_{jt} for any $i \neq j$ is r. We set $r = 0.5$ and consider the sample size $100, 300, 700$.

The factor model used as a shrinkage target is

$$X_{it}^2 = \mu_i + v_t \beta_{i0} + v_t^2 \beta_{i1} + \epsilon_{it},\ t = 1,2,\ldots,T \quad \text{and} \quad i = 1,2,\ldots,J. \tag{7}$$

So, in Eq. (1), $\mathbf{V_t} = [v_t\ v_t^2]'$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_{10}\ \beta_{20} \cdots \beta_{J0} \\ \beta_{11}\ \beta_{21} \cdots \beta_{J1} \end{bmatrix}'$.

3.1 GMM estimation and hypothesis testing

We are interested in estimating and testing hypotheses about the standard deviations, σ_i, of X_{it}, where $i = 1,2,\ldots,J$. To simplify the analysis, we assume that we know $E(X_{it}) = 0$ for all i and t. So the moment conditions for the GMM estimation of $\boldsymbol{\sigma} = [\sigma_1\ \ldots\ \sigma_J]'$ are

$$E[\mathbf{u}(\mathbf{X_t}, \boldsymbol{\sigma})] = E[X_{1t}^2 - \sigma_1^2\ \cdots\ X_{Jt}^2 - \sigma_J^2]' = \mathbf{0}_{J\times 1}. \tag{8}$$

This leads to the exactly identified GMM estimator

$$\hat{\boldsymbol{\sigma}} = \left[\left(\frac{1}{T} \sum_{t=1}^{T} X_{1t}^2 \right)^{1/2} \cdots \left(\frac{1}{T} \sum_{t=1}^{T} X_{Jt}^2 \right)^{1/2} \right]', \tag{9}$$

where $\hat{\boldsymbol{\sigma}} = [\hat{\sigma}_1\ \ldots\ \hat{\sigma}_J]'$.

The first hypothesis of interest for $\boldsymbol{\sigma} = [\sigma_1\ \ldots\ \sigma_J]'$ is the same as in Burnside and Eichenbaum (1996). The null is

$$H_0^F : \sigma_1 = \sigma_2 = \cdots = \sigma_J = 1$$

[3] We discuss the case where $\mathbf{X}_t$ follows a GARCH (1,1) process in Section 3.3.

with the alternative

$$H_1^F : \text{the null is wrong.}$$

We denote this hypothesis by a superscript 'F'.

The other hypothesis we consider does not contain full information about σ_i. The null is

$$H_0^P : \sigma_1 = \sigma_2 = \cdots = \sigma_J$$

with the alternative

$$H_1^P : \text{the null is wrong.}$$

The hypothesis H_0^P tests equality among σ_i without imposing any specific value on themselves.

The specific Wald statistic that we use to test H_0^F is given by

$$W_T^F = T(\hat{\boldsymbol{\sigma}} - \mathbf{1}_{J\times 1})'\mathbf{V}_T^{-1}(\hat{\boldsymbol{\sigma}} - \mathbf{1}_{J\times 1}). \tag{10}$$

Here $\mathbf{V}_T$ denotes a generic estimator of the asymptotic covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_{\mathbf{0}})$, where $\boldsymbol{\sigma}_{\mathbf{0}}$ is the true value of $\boldsymbol{\sigma}$. Given well-behaved estimators $\hat{\boldsymbol{\sigma}}$ and $\mathbf{V}_T$, $W_T^F \xrightarrow{d} \chi^2(J)$. Similarly, the specific Wald statistic for testing H_0^P is

$$W_T^P = T\hat{\boldsymbol{\sigma}}'\mathbf{A}'(\mathbf{A}\mathbf{V}_T\mathbf{A}')^{-1}\mathbf{A}\hat{\boldsymbol{\sigma}}, \tag{11}$$

where $\mathbf{A} = [-\mathbf{1}_{J-1\times 1}\ \mathbf{I}_{J-1\times J-1}]$. Also, we have $W_T^P \xrightarrow{d} \chi^2(J-1)$.

For the moment conditions, $E[\mathbf{u}(\mathbf{X}_t, \boldsymbol{\sigma})] = \mathbf{0}_{J\times 1}$, which are used to estimate $\boldsymbol{\sigma}$, the asymptotic covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_{\mathbf{0}})$ is given by $\mathbf{V}_0 = (\mathbf{D}_0'\mathbf{S}_0^{-1}\mathbf{D}_0)^{-1}$, where

$$\mathbf{D}_0 = E\frac{\partial \mathbf{u}(\mathbf{X}_t, \boldsymbol{\sigma}_{\mathbf{0}})}{\partial \boldsymbol{\sigma}'} \quad \text{and} \quad \mathbf{S}_0 = \sum_{j=-\infty}^{\infty} E\mathbf{u}(\mathbf{X}_t, \boldsymbol{\sigma}_{\mathbf{0}})\mathbf{u}(\mathbf{X}_{t-j}, \boldsymbol{\sigma}_{\mathbf{0}})'. \tag{12}$$

For the specified DGP, we can characterize $\mathbf{D}_0$ and $\mathbf{S}_0$ explicitly. First, $\mathbf{D}_0$ is given by

$$\mathbf{D}_0 = \begin{bmatrix} -2\sigma_{0,1} & 0 & \cdots & 0 \\ 0 & -2\sigma_{0,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -2\sigma_{0,J} \end{bmatrix} \tag{13}$$

where $(\sigma_{0,1}\ \sigma_{0,2}\ \cdots\ \sigma_{0,J})' = \boldsymbol{\sigma}_{\mathbf{0}} = \mathbf{1}_{J\times 1}$. Second, $\mathbf{S}_0$ is given by

$$\mathbf{S}_0 = \begin{bmatrix} \mathrm{Var}(X_{1t}^2) & \mathrm{Cov}(X_{1t}^2, X_{2t}^2) & \cdots & \mathrm{Cov}(X_{1t}^2, X_{Jt}^2) \\ \mathrm{Cov}(X_{2t}^2, X_{1t}^2) & \mathrm{Var}(X_{2t}^2) & \cdots & \mathrm{Cov}(X_{2t}^2, X_{Jt}^2) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{Cov}(X_{Jt}^2, X_{1t}^2) & \mathrm{Cov}(X_{Jt}^2, X_{2t}^2) & \cdots & \mathrm{Var}(X_{Jt}^2) \end{bmatrix} \tag{14}$$

where $\mathrm{Var}(X_{it}^2) = E(X_{it}^4) - [E(X_{it}^2)]^2 = 2$ since X_{it} follows a normal distribution, for $i = 1, 2, \ldots, J$. Moreover, $\mathrm{Cov}(X_{it}^2, X_{jt}^2) = E(X_{it}^2X_{jt}^2) - E(X_{it}^2)E(X_{jt}^2) =$

$2r^2$, since $E(X_{it}^2 X_{jt}^2) = E([\sqrt{r}v_t+\sqrt{1-r}e_{it}]^2[\sqrt{r}v_t+\sqrt{1-r}e_{jt}]^2) = r^2E(v_t^4)+2r(1-r)+(1-r)^2 = 2r^2+1$, for any $i \neq j$. Therefore, the true value of the asymptotic covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma_0})$ is given by

$$\mathbf{V}_0 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2}r^2 & \cdots & \frac{1}{2}r^2 \\ \frac{1}{2}r^2 & \frac{1}{2} & \cdots & \frac{1}{2}r^2 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{2}r^2 & \frac{1}{2}r^2 & \cdots & \frac{1}{2} \end{bmatrix}. \tag{15}$$

3.2 Estimators of the covariance matrix

In practice, we do not know the underlying DGP, so the true value of $\mathbf{V}_0$ cannot be observed. Here, we discuss various estimators of the asymptotic covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma_0})$. Accordingly, the corresponding estimator of $\mathbf{V}_0$ is given by $\mathbf{V}_T = (\mathbf{D}_T'\mathbf{S}_T^{-1}\mathbf{D}_T)^{-1}$, where $\mathbf{D}_T$ and $\mathbf{S}_T$ are the consistent estimators of $\mathbf{D}_0$ and $\mathbf{S}_0$.

By Newey and West (1987), a general nonparametric version of $\mathbf{S}_T$ can be written as

$$\mathbf{S}_T = \sum_{j=-(T-1)}^{T-1} k\left(\frac{j}{B_T}\right)\hat{\boldsymbol{\Omega}}_j, \tag{16}$$

where

$$\hat{\boldsymbol{\Omega}}_j = \begin{cases} \frac{1}{T}\sum\limits_{t=j+1}^{T} \mathbf{u}(\mathbf{X}_t, \hat{\boldsymbol{\sigma}})\mathbf{u}(\mathbf{X}_{t-j}, \hat{\boldsymbol{\sigma}})' & \text{for } j \geq 0 \\ \frac{1}{T}\sum\limits_{t=-j+1}^{T} \mathbf{u}(\mathbf{X}_{t+j}, \hat{\boldsymbol{\sigma}})\mathbf{u}(\mathbf{X}_t, \hat{\boldsymbol{\sigma}})' & \text{for } j < 0. \end{cases} \tag{17}$$

and

$$k(x) = \begin{cases} 1-|x| & \text{for } |x| \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

Here B_T is a scalar that determines the bandwidth of the lag window, $k(\cdot)$. This estimator does not exploit any of the underlying information. To simplify our analysis, we set $B_T < 1$ based on the assumption that the time series data are serially and independently generated. This implies that the estimator has an ijth element given by $\frac{1}{T}\sum\limits_{t=1}^{T}(X_{it}^2 - \hat{\sigma}_i^2)(X_{jt}^2 - \hat{\sigma}_j^2)$. This estimator is denoted by $\mathbf{S}_T^{(1)}$, and it serves as the first estimator of $\mathbf{S}_0$. From $\mathbf{u}(\mathbf{X}_t, \boldsymbol{\sigma})$, the estimator of $\mathbf{D}_0$ is straightforwardly given by $\mathbf{D}_T^{(1)}$, which is a diagonal matrix with the iith element equal to $-2\hat{\sigma}_i$.

The second estimator, $\mathbf{S}_T^{(2)}$, exploits the fact that X_{it} is Gaussian, which implies that $E(X_{it}^4) = 3\sigma_i^4$. So $\mathbf{S}_T^{(2)}$ is a matrix with the iith element given by $2\hat{\sigma}_i^4$ and with the same nondiagonal elements as in $\mathbf{S}_T^{(1)}$. That is to say, the estimators of the diagonal elements are obtained by estimating the parameter σ and the nondiagonal elements are estimated by the nonparametric method.

We can impose the null hypothesis H_0^F, where $\sigma_i = 1$ for $i = 1, 2, \ldots, J$. This suggests a third estimator, $\mathbf{S}_T^{(3)}$, which has the iith element equal to 2 and the same nondiagonal elements as in $\mathbf{S}_T^{(2)}$. Similarly, the null hypothesis can be imposed on $\mathbf{D}_T^{(1)}$ to yield another estimator of $\mathbf{D}_0$, i.e, $\mathbf{D}_T^{(2)}$, which is a diagonal matrix with the iith element equal to -2 for $i = 1, 2, \ldots, J$.

Corresponding to each of these estimators of $\mathbf{S}_0$, the first three estimators of $\mathbf{V}_0$ are given by

$$\mathbf{V}_T^{(k)} = [\mathbf{D}_T^{(1)'} (\mathbf{S}_T^{(k)})^{-1} \mathbf{D}_T^{(1)}]^{-1}, \; k = 1, 2, 3. \tag{19}$$

In addition, we also consider the estimator

$$\mathbf{V}_T^{(4)} = [\mathbf{D}_T^{(2)'} (\mathbf{S}_T^{(3)})^{-1} \mathbf{D}_T^{(2)}]^{-1}. \tag{20}$$

All these estimators correspond to those in panels (d), (f), (g) and (h) of Table 1 in Burnside and Eichenbaum (1996). In this paper, we assume that any information about the nondiagonal elements of $\mathbf{S}_0$ cannot be explored, whereas Burnside and Eichenbaum (1996) impose a diagonal matrix on $\mathbf{S}_T$ since X_{it} are mutually and independently generated. This ensures our results are comparable to those of Burnside and Eichenbaum (1996) which employ almost full information about the underlying DGP and the null hypothesis to improve the finite-sample performance of the Wald tests. However, the practitioner usually does not have full information about the model, so our settings are more realistic. In addition, this comparison allows us to test whether imposing on $\mathbf{S}_T$ partial information from the underlying DGP and the null hypothesis can improve the finite-sample performance of the Wald tests.

Based on the findings of Burnside and Eichenbaum (1996), we find that the anomalies associated with the finite-sample distribution of Wald statistics are closely related to the finite-sample distribution of $\mathbf{V}_T$ in our settings. In order to justify this conjecture, we set $\mathbf{V}_T^{(5)} = \mathbf{V}_0$.

The last estimator of $\mathbf{V}_0$ is based on the shrinkage estimator of $\mathbf{S}_0$, or

$$\mathbf{V}_T^{(6)} = [\mathbf{D}_T^{(1)'} (\mathbf{S}_T^{(s)})^{-1} \mathbf{D}_T^{(1)}]^{-1}. \tag{21}$$

For hypothesis H_0^P, the information from the null is implemented in a manner which suggests $\sigma_1 = \sigma_2 = \ldots = \sigma_J = \frac{1}{J}\sum_{j=1}^{J} \hat{\sigma_j}$. This indicates that $\mathbf{S}_T^{(3*)}$ is a matrix with the iith element equal to $2\frac{1}{J}\sum_{j=1}^{J} \hat{\sigma_j}$ for all i and with the same nondiagonal elements as in $\mathbf{S}_T^{(2)}$, and that $\mathbf{D}_T^{(2*)}$ is a diagonal matrix with the iith element equal to $-2\frac{1}{J}\sum_{j=1}^{J} \hat{\sigma_j}$ for all i. From this, six different estimators of $\mathbf{V}_0$ can be created, as before.

In addition, we discuss whether incorrect information from the underlying DGP will distort the size of the Wald tests for H_0^P by imposing a diagonal matrix on $\mathbf{S}_T$.

3.3 Monte Carlo experiments

We use MATLAB to simulate $10,000$ sets of $\{\mathbf{X}_t\}_{t=1}^T$, where $\mathbf{X}_t=[X_{1t} \cdots X_{Jt}]'$. Following Burnside and Eichenbaum (1996), we set $J = 20$. For each data set, we estimate six different estimators of the covariance matrix and conduct the Wald test for H_0^F and H_0^P, denoting W_T^F and W_T^P as the corresponding Wald test statistics, respectively.

Table 1 summarizes the results of H_0^F for this simulation with T=100, 300, 700. The panels correspond to the different estimators of $\mathbf{V}_0$. The first column of each panel is the asymptotic size of the tests. The entries in the other columns are the rejection frequencies of the Wald test corresponding to the different sizes of the test. This table reveals four findings. First, the Wald test performs poorly in finite samples. From Panel (a) of Table 1, the finite-sample sizes of the test exceed the nominal levels. This result is similar to that obtained by Burnside and Eichenbaum (1996). Although the finite-sample sizes are improved as T increases from 100 to 700, the deviation still exists.

Second, partially imposing restrictions from the underlying DGP or the null hypothesis does not improve the finite-sample performance of the Wald test significantly. For $T = 100$, the finite-sample sizes of the test are improved by imposing the information from the DGP and the null hypothesis; however, for $T = 300$ and 700, the sizes of the test are not improved, and are even less accurate, as indicated in panels (b)–(d) of Table 1.

Third, as Panel (e) shows, the finite-sample size of the test is close to the corresponding nominal levels when the Wald statistics are calculated by using a true value of $\mathbf{V}_0$. These findings are consistent with those showed by Burnside and Eichenbaum (1996).

Finally, Panel (f) indicates that the shrinkage method can improve the finite-sample performance of the Wald test, compared with panels (a)-(d). The improvement is most prominent for $T = 300$. When $T = 300$, the rejection frequencies in Panel (f) are just half of those in the first four panels, and are closer to the nominal levels.

Since H_0^P does not contain enough information about $\mathbf{S}_T$, we conduct two experiments. In the first one, information from the null is implemented in such a way that suggests $\sigma_1 = \sigma_2 = \ldots = \sigma_J = \frac{1}{J}\sum_{j=1}^J \hat{\sigma_j}$. In the second one, incorrect information -that X_{it} is mutually independent- is imposed directly on the estimation of $\mathbf{S}_0$. We then follow the same steps as in Table 1 and obtain the rejection frequencies of the Wald tests.

The left-hand panels of Table 2 summarizes the results for the first experiment. We see that these results are quite similar to Table 1. This is not surprising because $\frac{1}{J}\sum_{j=1}^J \hat{\sigma_j}$ is close to the true value, 1. However, if we look at the right-hand panels of Table 2, as panels (a′)–(d′) indicate, the finite-sample sizes are very close to zero. The larger the sample size, T, the closer the finite-sample sizes are to zero. This means that the restriction from such incorrect information distorts the sizes of the Wald test. In contrast, Panel (f′) has finite-sample sizes similar to those in Panel (f), implying that the improve-

Table 1 Finite-Sample Performance of Wald Test for $H_0^F : \sigma_1 = \sigma_2 = \cdots = \sigma_J = 1$. The data are simulated by Gaussian distributions with a mutual correlation of 0.5. $\mathbf{V}_T$ is the estimated covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma_0})$, and $\mathbf{V}_T = (\mathbf{D}_T'\mathbf{S}_T^{-1}\mathbf{D}_T)^{-1}$. $\mathbf{S}_T^{(1)}$ is the Newey-West estimator with no lags. $\mathbf{S}_T^{(2)}$ is a matrix with the iith element given by $2\hat{\sigma}_i^4$ and the same nondiagonal elements as in $\mathbf{S}_T^{(1)}$. $\mathbf{S}_T^{(3)}$ has the iith element equal to 2 and the same nondiagonal elements as in $\mathbf{S}_T^{(2)}$. $\mathbf{S}_T^{(s)}$ is the shrinkage estimator. $\mathbf{D}_T^{(1)}$ is a diagonal matrix with the iith element equal to $-2\hat{\sigma}_i$. $\mathbf{D}_T^{(2)}$ is a diagonal matrix with the iith element equal to -2.

	Finite-sample size (%)		
Asymptotic size	T=100	T=300	T=700
	(a) $\mathbf{S}_T^{(1)}$ and $\mathbf{D}_T^{(1)}$		
1%	27.65	6.01	2.52
5%	45.54	15.35	9.27
10%	54.49	25.26	15.28
	(b) $\mathbf{S}_T^{(2)}$ and $\mathbf{D}_T^{(1)}$		
1%	26.32	6.64	2.64
5%	39.05	15.43	8.90
10%	45.36	25.18	14.91
	(c) $\mathbf{S}_T^{(3)}$ and $\mathbf{D}_T^{(1)}$		
1%	20.30	9.38	3.12
5%	28.55	18.22	9.58
10%	34.52	26.71	15.92
	(d) $\mathbf{S}_T^{(3)}$ and $\mathbf{D}_T^{(2)}$		
1%	20.34	7.98	2.53
5%	30.37	16.89	8.97
10%	36.94	26.29	15.16
	(e) True value of $\mathbf{V}_0$		
1%	0.94	0.84	0.96
5%	4.87	5.11	5.13
10%	10.32	10.70	9.78
	(f) $\mathbf{S}_T^{(s)}$ and $\mathbf{D}_T^{(1)}$		
1%	11.00	3.07	1.72
5%	22.65	9.96	7.17
10%	32.17	17.99	12.45

ment from the shrinkage estimator is robust to the change of the shrinkage target.

Many economic variables demonstrate conditional heteroscedasticity. In order to explore how this dependence affects our approach, we use a GARCH (1,1) model to simulate the data and repeat the previous experiments. The data generating process is

$$X_{it} = \sqrt{r}v_t + \sqrt{1-r}e_{it} \text{ for all } i \text{ and } t,$$

Table 2 Finite-Sample Performance of Wald Test for $H_0^P : \sigma_1 = \sigma_2 = \cdots = \sigma_J$. The data are simulated by Gaussian distributions with a mutual correlation of 0.5. $\mathbf{V}_T$ is the estimated covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_\mathbf{0})$, and $\mathbf{V}_T = (\mathbf{D}_T'\mathbf{S}_T^{-1}\mathbf{D}_T)^{-1}$. $\mathbf{S}_T^{(1)}$ is the Newey-West estimator with no lags. $\mathbf{S}_T^{(2)}$ is a matrix with the iith element given by $2\hat{\sigma}_i^4$ and with the same nondiagonal elements as in $\mathbf{S}_T^{(1)}$. $\mathbf{S}_T^{(3*)}$ has the iith element equal to $2\frac{1}{J}\sum_{j=1}^J \hat{\sigma_j}$ and the same nondiagonal elements as in $\mathbf{S}_T^{(2)}$. $\mathbf{S}_T^{(s)}$ is the shrinkage estimator defined in Eq. (4) while $\mathbf{S}_T^{(s*)}$ is the shrinkage estimator when X_{it} is assumed to be i.i.d. (or, $\boldsymbol{\beta} = \mathbf{0}$). $\mathbf{D}_T^{(1)}$ is a diagonal matrix with the iith element equal to $-2\hat{\sigma}_i$. $\mathbf{D}_T^{(2*)}$ is a diagonal matrix with the iith element equal to $-2\frac{1}{J}\sum_{j=1}^J \hat{\sigma_j}$. The right-hand side panels report the results when we assume $\mathbf{S}_T$ is a diagonal matrix and keep all other settings unchanged.

Asymptotic size	Finite-sample size (%) T=100	T=300	T=700	Asymptotic size	Finite-sample size (%) T=100	T=300	T=700
	(a) $\mathbf{S}_T^{(1)}$ and $\mathbf{D}_T^{(1)}$				(a') Diagonal $\mathbf{S}_T^{(1)}$, $\mathbf{D}_T^{(1)}$		
1%	17.03	3.31	1.72	1%	0.31	0.05	0.02
5%	33.09	11.31	7.66	5%	1.81	0.62	0.37
10%	44.06	20.43	13.58	10%	3.57	1.64	1.14
	(b) $\mathbf{S}_T^{(2)}$ and $\mathbf{D}_T^{(1)}$				(b') Diagonal $\mathbf{S}_T^{(2)}$, $\mathbf{D}_T^{(1)}$		
1%	18.93	4.21	1.91	1%	0.10	0.02	0.01
5%	29.77	11.54	7.34	5%	0.60	0.33	0.32
10%	37.48	20.56	13.20	10%	1.77	1.22	0.92
	(c) $\mathbf{S}_T^{(3*)}$ and $\mathbf{D}_T^{(1)}$				(c') Diagonal $\mathbf{S}_T^{(3*)}$, $\mathbf{D}_T^{(1)}$		
1%	22.79	7.07	2.53	1%	0.07	0.03	0.00
5%	31.87	15.49	8.95	5%	0.52	0.29	0.31
10%	37.74	24.11	14.83	10%	1.61	1.07	0.87
	(d) $\mathbf{S}_T^{(3*)}$ and $\mathbf{D}_T^{(2*)}$				(d') Diagonal $\mathbf{S}_T^{(3*)}$, $\mathbf{D}_T^{(2*)}$		
1%	17.43	4.39	1.75	1%	0.03	0.03	0.00
5%	26.48	11.74	7.44	5%	0.42	0.25	0.29
10%	32.35	20.09	12.87	10%	1.40	1.06	0.86
	(e) True value of $\mathbf{V}_0$				(e') True value of $\mathbf{V}_0$		
1%	1.06	0.83	0.95	1%	1.06	0.83	0.95
5%	5.19	5.08	5.02	5%	5.19	5.08	5.02
10%	9.79	10.64	9.86	10%	9.79	10.64	9.86
	(f) $\mathbf{S}_T^{(s)}$ and $\mathbf{D}_T^{(1)}$				(f') $\mathbf{S}_T^{(s*)}$, $\mathbf{D}_T^{(1)}$		
1%	3.63	1.26	1.07	1%	3.57	1.52	1.36
5%	12.49	6.92	5.77	5%	12.53	7.52	6.47
10%	20.52	13.40	10.78	10%	19.40	13.77	11.44

where v_t is i.i.d. and $N(0,1)$ for all t and for each i. Here, $e_{it} = \sqrt{h_{it}}z_t$,

$$z_t \sim N(0,1) \text{ and } h_{it} = \alpha_i + \beta_i e_{i,t-1}^2 + \gamma_i h_{i,t-1}.$$

We set $\alpha_i = 0.5$, $\beta_i = 0.25$ and $\gamma_i = 0.25$ for all i. As before, we set r to be 0.5. We use the optimal bandwidth, B_T, to obtain the Newey-West estimators, denoted by $\mathbf{S}_T^{(1)}$. As for the shrinkage estimator, we still treat the data as if they are i.i.d. The results are reported in Table 3. We find that when $\{X_{it}, t \in Z\}$ follows a GARCH (1,1) model, the performance of the shrinkage method is still better than of the alternative methods when $T = 100, 300$. When $T = 700$, the performance of the shrinkage estimator is comparable to those alternatives. We can conservatively conclude that if the true DGP is a GARCH (1,1) model, then the shrinkage method, at a minimum, does not adversely affect our results.

In summary, the results from the experiments imply that partial information from the underlying DGP and from the null hypothesis could not improve the results of the finite-sample sizes of the Wald test, and that imposing incorrect information leads to a distortion of the sizes of the Wald test. However, the shrinkage estimator of the covariance matrix significantly improves the finite-sample properties of the Wald test under our simple Gaussian DGP. This improvement is robust to the choice of the shrinkage target.

4 Empirical Application

Mean-variance efficiency is a key issue in finance. Researchers try to construct portfolios which are mean-variance efficient. A portfolio is deemed to be mean-variance efficient if either of two conditions are met. The first, given the level of the expected portfolio return, has the smallest variance compared with the other portfolios, the second, given the level of variance of the return, has the highest mean. These comparisons are mostly conducted based on the observations of the sample mean and the sample variance. We can get more reliable inferences from a Wald test. In this section, we illustrate the importance of our method by studying mean-variance efficiency properties.

We collect the monthly returns of the Fama-French 25 portfolios from July 1963 to December 1990 (330 observations). [4] These portfolios are formed on the basis of their sizes (market equity) and the ratios of book equity to market equity for all stocks listed on the NYSE, the AMEX, and the NASDAQ. We treat them as 25 different risky assets. By holding these assets with J different weights, we can obtain J different portfolios.

In the first step, we form only one portfolio, which is mean-variance efficient. According to Markowitz (1952), given the mean of the return, we minimize the variance of the return of this new portfolio as

$$\min_{\mathbf{w}} \mathbf{w}' \boldsymbol{\Sigma} \mathbf{w} \tag{22}$$

subject to

$$\mathbf{w}'\mathbf{1} = 1 \text{ and } \mathbf{w}'\boldsymbol{\mu} = q,$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of the asset returns, $\boldsymbol{\mu}$ is the mean vector of the returns, $\mathbf{w}$ is the weight, q is the mean of the return for this portfolio and $\mathbf{1}$ is a vector of 1.

[4] URL: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Table 3 Finite-Sample Performance of Wald Test for $H_0^F : \sigma_1 = \sigma_2 = \cdots = \sigma_J = 1$ and $H_0^P : \sigma_1 = \sigma_2 = \cdots = \sigma_J$. The data are simulated by a GARCH (1,1) model. $\mathbf{V}_T$ is the estimated covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_\mathbf{0})$, and $\mathbf{V}_T = (\mathbf{D}_T' \mathbf{S}_T^{-1} \mathbf{D}_T)^{-1}$. $\mathbf{S}_T^{(1)}$ is the Newey-West estimator with optimal bandwidth. $\mathbf{S}_T^{(2)}$ is a matrix with the iith element given by $2\hat{\sigma}_i^4$ and with the same nondiagonal elements as in $\mathbf{S}_T^{(1)}$. $\mathbf{S}_T^{(3)}$ has the iith element equal to 2 and with the same nondiagonal elements as in $\mathbf{S}_T^{(2)}$. $\mathbf{S}_T^{(3*)}$ has the iith element equal to $2\frac{1}{J}\sum_{j=1}^{J}\hat{\sigma_j}$ and the same nondiagonal elements as in $\mathbf{S}_T^{(2)}$. $\mathbf{S}_T^{(s)}$ is the shrinkage estimator. $\mathbf{D}_T^{(1)}$ is a diagonal matrix with the iith element equal to $-2\hat{\sigma}_i$. $\mathbf{D}_T^{(2)}$ is a diagonal matrix with the iith element equal to -2. $\mathbf{D}_T^{(2*)}$ is a diagonal matrix with the iith element equal to $-2\frac{1}{J}\sum_{j=1}^{J}\hat{\sigma_j}$.

$H_0 : \sigma_i = 1$ for all i				$H_0 : \sigma_i = \sigma_j$ for all $i \neq j$.			
	Finite-sample size (%)				Finite-sample size (%)		
Asymptotic size	T=100	T=300	T=700	Asymptotic size	T=100	T=300	T=700
(a) $\mathbf{S}_T^{(1)}$ and $\mathbf{D}_T^{(1)}$				(a) $\mathbf{S}_T^{(1)}$ and $\mathbf{D}_T^{(1)}$			
1%	72.18	29.10	13.14	1%	64.62	24.90	11.68
5%	83.48	47.64	30.08	5%	78.32	43.06	27.18
10%	87.92	58.62	40.50	10%	84.34	53.91	37.96
(b) $\mathbf{S}_T^{(2)}$ and $\mathbf{D}_T^{(1)}$				(b) $\mathbf{S}_T^{(2)}$ and $\mathbf{D}_T^{(1)}$			
1%	51.48	50.08	47.44	1%	52.18	48.38	45.04
5%	64.50	62.10	55.01	5%	63.70	60.91	52.44
10%	73.18	68.34	59.06	10%	71.66	67.24	56.84
(c) $\mathbf{S}_T^{(3)}$ and $\mathbf{D}_T^{(1)}$				(c) $\mathbf{S}_T^{(3*)}$ and $\mathbf{D}_T^{(1)}$			
1%	47.34	44.28	32.02	1%	48.08	46.08	34.62
5%	62.50	55.46	39.58	5%	63.10	56.78	41.24
10%	69.42	60.76	43.54	10%	70.40	62.72	44.82
(d) $\mathbf{S}_T^{(3)}$ and $\mathbf{D}_T^{(2)}$				(d) $\mathbf{S}_T^{(3*)}$ and $\mathbf{D}_T^{(2*)}$			
1%	47.48	45.02	33.94	1%	46.50	44.21	32.86
5%	62.86	56.32	42.66	5%	61.91	55.48	40.38
10%	70.20	62.38	47.00	10%	70.16	62.01	44.80
(e) True value of $\mathbf{V}_0$				(e) True value of $\mathbf{V}_0$			
1%	3.46	3.08	2.91	1%	3.47	3.09	2.90
5%	9.78	9.10	8.94	5%	10.00	9.33	9.03
10%	16.73	15.99	14.61	10%	16.76	16.09	14.54
(f) $\mathbf{S}_T^{(s)}$ and $\mathbf{D}_T^{(1)}$				(f) $\mathbf{S}_T^{(s)}$ and $\mathbf{D}_T^{(1)}$			
1%	33.14	20.60	14.82	1%	24.28	17.26	13.71
5%	52.20	35.06	32.67	5%	43.92	34.41	30.48
10%	63.02	47.54	40.63	10%	56.54	45.22	37.48

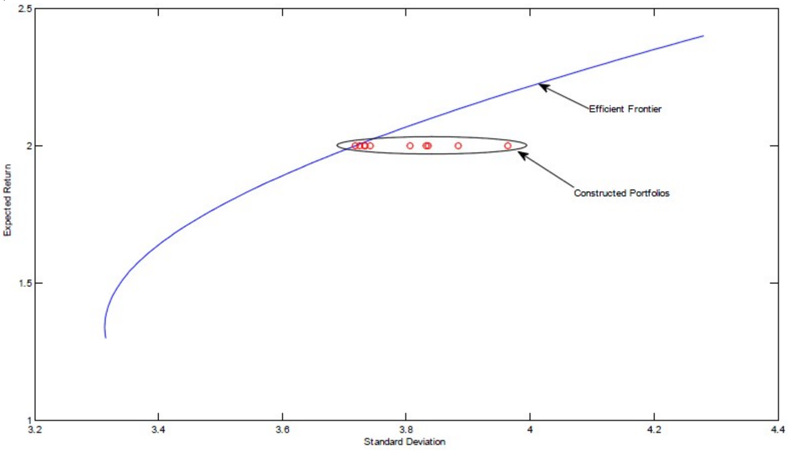
Then, a frontier portfolio can be obtained by choosing

$$\mathbf{w} = \frac{C - qB}{AC - B^2}\boldsymbol{\Sigma}^{-1}\mathbf{1} + \frac{qA - B}{AC - B^2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \tag{23}$$

where $A = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}$, $B = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ and $C = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$.

In the second step, we can create other $J - 1$ portfolios, which have the same q but a different sample variance. Each of these portfolios is constructed by randomly generating the first 23 elements of $\bar{\mathbf{w}}$ from $\mathbf{w}+U[-1/27.5, 1/27.5]$ and then calculating the remaining two elements of $\bar{\mathbf{w}}$ according to two restrictions: $\bar{\mathbf{w}}'\mathbf{1} = 1$ and $\bar{\mathbf{w}}'\boldsymbol{\mu} = q$. A rough figure is plotted in Figure 1.

Fig. 1 Mean-Variance Efficienct Frontier. The curve is the mean-variance efficient frontier. Small circles are the portfolios when $q = 2\%$ and $J = 10$. One of them is located on the curve, and the others are below the curve.



The curve in Figure 1 is the mean-variance efficient frontier. Any point on this frontier indicates the minimum standard deviation a portfolio can achieve given the mean (e.g., the value of q). If we focus on the case where $q = 2\%$ and $J = 10$, then the small circles are the mean and the standard deviations of the portfolios we construct. One of them is mean-variance efficient, located on the efficient frontier; the others are located below the curve.

Then, we test the null hypothesis that

$$H_0 : \sigma_1 = \sigma_2 = \cdots = \sigma_J.$$

We conduct two kinds of Wald tests. The first uses the sample analog to estimate the covariance matrix, and the second uses the Fama-French three-factor model to derive the shrinkage estimator. The Fama-French three-factor model is:

$$R_i = R_f + \beta_{1i}(R_m - R_f) + \beta_{2i}\text{SMB} + \beta_{3i}\text{HML} + e_i,$$

where R_i is the portfolio return, R_f is the risk-free return, R_m is the market portfolio return and SMB is the average return of the three small portfolios (based on market equity) minus the average return of the three big portfolios. HML is the average return of the two value portfolios (based on the ratios of book equity to market equity) minus the average return of the two growth portfolios.

We try different combinations of q and J. The results are summarized in Tables 4 and 5. We see that these two tables deliver quite different results. Table 4 suggests that we should reject the null hypothesis for any value of q and J, while Table 5 leads us to the opposite conclusion. This implies that, in practice, the shrinkage method really matters. If we follow the conventional Wald test, we would believe that, given q, the frontier portfolio based on Eq. (23) is mean-variance efficient and unique since Eq. (23) suggests that there is one-to-one correspondence between q and $\mathbf{w}$. However, if we apply the shrinkage method to the estimation of the covariance matrix, the testing results suggest that we can find yet another set of portfolios which has the same mean and the same variance with the frontier portfolio. The p-values that are marginally non-significant (q=2% in Table 5) should be treated with caution, since the test is slightly liberal.

Table 4 W_T^J statistics when the weighting matrix is estimated by the sample analogue. $p-$values are in the parentheses.

$q \setminus J$	5	6	7	8	9	10
2.00%	55.16(0.00)	57.40(0.00)	60.11(0.00)	63.62(0.00)	64.11(0.00)	69.43(0.00)
2.05%	54.91(0.00)	57.25(0.00)	58.58(0.00)	63.57(0.00)	64.13(0.00)	69.45(0.00)
2.10%	54.64(0.00)	57.09(0.00)	58.44(0.00)	63.51(0.00)	64.13(0.00)	69.44(0.00)
2.15%	54.36(0.00)	56.92(0.00)	58.29(0.00)	63.44(0.00)	64.12(0.00)	69.44(0.00)
2.20%	54.06(0.00)	56.74(0.00)	58.13(0.00)	63.36(0.00)	64.11(0.00)	69.41(0.00)
2.25%	53.75(0.00)	56.54(0.00)	57.95(0.00)	63.28(0.00)	64.08(0.00)	69.37(0.00)
2.30%	53.42(0.00)	56.33(0.00)	57.76(0.00)	63.18(0.00)	64.04(0.00)	69.31(0.00)
2.35%	53.08(0.00)	56.12(0.00)	57.57(0.00)	63.08(0.00)	64.00(0.00)	69.24(0.00)
2.40%	52.73(0.00)	55.89(0.00)	57.36(0.00)	62.97(0.00)	63.94(0.00)	69.16(0.00)
2.45%	52.36(0.00)	55.65(0.00)	57.15(0.00)	62.85(0.00)	63.87(0.00)	69.06(0.00)
2.50%	51.99(0.00)	55.40(0.00)	56.92(0.00)	62.73(0.00)	63.79(0.00)	68.95(0.00)

5 Conclusion

We propose using shrinkage method to remedy the small-sample properties of GMM-based Wald tests. The shrinkage method provides a more reliable estimator for the covariance matrix used in the Wald test, and hence improves the size of the Wald test. Simulations reveal that our method is superior to those suggested in Burnside and Eichenbaum (1996). Furthermore, we apply

Table 5 W_T^J statistics when the weighting matrix is estimated by the shrinkage method. $p-$values are in the parentheses.

$q \setminus J$	5	6	7	8	9	10
2.00%	7.96(0.09)	10.12(0.07)	11.60(0.07)	11.39(0.12)	11.65(0.17)	12.83(0.17)
2.05%	7.22(0.12)	9.21(0.10)	10.55(0.10)	10.37(0.17)	10.62(0.22)	11.70(0.23)
2.10%	6.54(0.16)	8.36(0.14)	9.58(0.14)	9.42(0.22)	9.66(0.29)	10.66(0.30)
2.15%	5.92(0.20)	7.58(0.18)	8.70(0.19)	8.56(0.29)	8.78(0.36)	9.70(0.38)
2.20%	5.36(0.25)	6.88(0.23)	7.89(0.25)	7.78(0.35)	7.98(0.44)	8.83(0.45)
2.25%	4.86(0.30)	6.24(0.28)	7.16(0.31)	7.08(0.42)	7.26(0.51)	8.04(0.53)
2.30%	4.41(0.35)	5.67(0.34)	6.51(0.37)	6.44(0.49)	6.61(0.58)	7.33(0.60)
2.35%	4.01(0.40)	5.16(0.40)	5.92(0.43)	5.87(0.55)	6.03(0.64)	6.68(0.67)
2.40%	3.65(0.46)	4.70(0.45)	5.38(0.50)	5.35(0.62)	5.50(0.70)	6.10(0.73)
2.45%	3.33(0.50)	4.28(0.51)	4.91(0.56)	4.90(0.67)	5.03(0.75)	5.58(0.78)
2.50%	3.04(0.55)	3.91(0.56)	4.49(0.61)	4.48(0.73)	4.60(0.80)	5.11(0.82)

the shrinkage method to mean-variance efficient portfolios and find that the conventional Wald test can be misleading when determining mean-variance efficiency, and that our method produces more reliable inferences.

References

1. Andersen T, Sørensen B (1996) GMM estimation of a stochastic volatility model: A Monte Carlo study. Journal of Business & Economic Statistics 14: 328-352.
2. Bekaert G, Hodrick R (2001) Expectations hypotheses tests. The Journal of Finance 56: 1357-1394.
3. Burnside C, Eichenbaum M (1996) Small-sample properties of GMM-based Wald tests. Journal of Business & Economic Statistics 7: 265-296.
4. Christiano L, Den Haan W (1996) Small-sample properties of GMM for business-cycle analysis. Journal of Business & Economic Statistics 14: 309-327.
5. Hansen LP (1982) Large sample properties of generalized method of moments estimators. Econometrica 50: 1029-1054.
6. Hansen L, Heaton J, Yaron A (1996) Finite-sample properties of some alternative GMM estimators. Journal of Business & Economic Statistics 14: 262-280.
7. Hall P, Horowitz L (1996) Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64: 891-916.
8. Kan R, Zhang C (1999) GMM tests of stochastic discount factor models with useless factors. Journal of Financial Economics 54: 103-127.
9. Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance 10: 603-621.
10. Newey W, West K (1987) A simple, positive semi-definite, heteroskedasticity and auto-correlation consistent covariance matrix. Econometrica 55: 703-08.
11. Ren Y, Shimotsu K (2009) Improvement in finite sample properties of the Hansen-Jagannathan distance test. Journal of Empirical Finance 16: 483-506.
12. Rothenberg T (1984) Approximating the distributions of econometric estimators and test statistics. Handbook of Econometrics 2: 881-935.