TECHNICAL WORKING PAPER SERIES

SMALL SAMPLE PROPERTIES
OF GENERALIZED METHOD OF
MOMENTS BASED WALD TESTS

Craig Burnside
Martin Eichenbaum

Technical Working Paper No. 155

SMALL SAMPLE PROPERTIES
OF GENERALIZED METHOD OF
MOMENTS BASED WALD TESTS

ABSTRACT

This paper assesses the small sample properties of Generalized Method of Moments (GMM) based Wald statistics. The analysis is conducted assuming that the data generating process corresponds to (i) a simple vector white noise process and (ii) an equilibrium business cycle model. Our key findings are that the small sample size of the Wald tests exceeds their asymptotic size, and that their size increases uniformly with the dimensionality of joint hypotheses. For tests involving even moderate numbers of moment restrictions, the small sample size of the tests greatly exceeds their asymptotic size. Relying on asymptotic distribution theory leads one to reject joint hypothesis tests *far* too often. We argue that the source of the problem is the difficulty of estimating the spectral density matrix of the GMM residuals, which is needed to conduct inference in a GMM environment. Imposing restrictions implied by the underlying economic model being investigated or the null hypothesis being tested on this spectral density matrix can lead to substantial improvements in the small sample properties of the Wald tests.

Craig Burnside
Department of Economics
University of Pittsburgh
Pittsburgh, PA 15260

Martin Eichenbaum
Department of Economics
Northwestern University
Evanston, IL 60208
and NBER

1 Introduction

This paper assesses the small sample properties of Generalized Method of Moments (GMM) based Wald statistics. The analysis is conducted assuming that the data generating process corresponds to (i) a simple vector white noise process and (ii) the equilibrium business cycle model considered in Burnside and Eichenbaum (1994). Our key findings are that the small sample size of the Wald tests exceeds their asymptotic size, and that their size increases uniformly with the dimensionality of joint hypotheses. For tests involving even moderate numbers of moment restrictions, the small sample size of the tests greatly exceeds their asymptotic size. Relying on asymptotic distribution theory leads one to reject joint hypothesis tests *far* too often. We argue that the source of the problem is the difficulty of estimating the spectral density matrix of the GMM residuals, which is needed to conduct inference in a GMM environment. Imposing restrictions implied by the underlying economic model being investigated or the null hypothesis being tested on this spectral density matrix can lead to substantial improvements in the small sample properties of the Wald tests.

A common approach to evaluating quantitative equilibrium business cycle models is to compare model and non-model based estimates of the second moments of aggregate time series. No uniform method for making these comparisons has emerged. Many authors in the Real Business Cycle (RBC) literature make these comparisons in a way that abstracts from sampling uncertainty in estimates of models' structural parameters (see for example Kydland and Prescott (1982) or Hansen (1985)). Other authors have estimated and tested RBC models using full information maximum likelihood methods (see for example Altug (1989), Christiano (1988), McGratten, Rogerson and Wright (1993) and Leeper and Sims (1994)).

An intermediate strategy is to simultaneously estimate model parameters and second moments of the data using a variant of Hansen's (1982) Generalized Method of Moments (GMM) procedure. Christiano and Eichenbaum (1992) show how, in this framework, simple Wald-type tests can be used to test models' implications for second moments of the data. Three advantages of this approach are that (i) at the estimation stage of the analysis one need not completely specify agents' environments, (ii) it is easy to specify which aspects of the data one wishes to concentrate on for diagnostic purposes, and (iii)

1

it is substantially less demanding from a computational point of view than maximum likelihood approaches. Use of this procedure has become more widespread. However its properties in small samples are not well understood. This is disturbing in light of recent results in the literature casting doubt on the extent to which asymptotic distribution theory provides a good approximation to various aspects of the small sample behavior of GMM based estimators.[1]

In this paper we address four basic questions concerning the performance of GMM based Wald statistics. First, does the small sample size of these tests closely approximate their asymptotic size? Second, do joint tests of several restrictions perform as well or worse than tests of simple hypotheses? Third, how can modeling assumptions, or restrictions imposed by hypotheses themselves, be used to improve the performance of these tests? Fourth, what practical advice, if any, can be given to the practitioner?

We answer these questions under two assumptions about the data generating process. First, we assume that the true process generating the macro time series is the equilibrium business cycle model developed in Burnside and Eichenbaum (1994). This case is of interest for two reasons: (i) the model generates time series that in several respects resemble U.S. data, and (ii) we can study issues of size and inference in an applied context. Second, we assume that the data generating process corresponds to Gaussian vector white noise. Working with such a simple process allows us to assess whether the findings that emerge with the more complicated data process also arise in simpler environments. In addition we find it easier to build intuition about our results in the simpler environment.

Our main findings can be summarized as follows. First, there is a strong tendency for GMM based Wald tests to over-reject. Second, the small sample size of these tests increases uniformly as the dimension of joint tests increases. For even moderate number of restrictions, the small sample size is dramatically larger than the asymptotic size of the test. Indeed correcting for the small sample properties of the Wald test turns out to have a substantive impact on inference about the empirical performance of the equilibrium business cycle model that is being analyzed. Third, the basic problem is difficulty in accurately estimating the spectral density matrix of the GMM error terms. We investigate various nonparametric estimators of this matrix that have been suggested in the literature. While

---

[1]See for Tauchen (1986), Kocherlakota (1990), Ferson and Foerster (1991), Burnside (1992), Fuhrer, Moore and Schuh (1993), Neely (1993), Christiano and den Haan (1994) and West and Wilcox (1994).

there is there is some sensitivity to which nonparametric estimator is used, these differences do not affect our basic conclusions. Fourth, we argue that the size characteristics of the Wald tests can be improved if the analyst imposes restrictions that emerge from the model or the hypothesis being tested when estimating the covariance matrix component of the Wald statistic. Not only does such information improve the size of simple tests, it significantly ameliorates the problems associated with tests of joint hypotheses.

The remainder of this paper is organized as follows. Section 2 considers the case of the Gaussian white noise generating process. In Section 3 we discuss the case where the data are generated from an equilibrium business cycle model. Section 4 contains some concluding remarks.

2 Gaussian White Noise Data Generating Processes

In this section we consider the small sample properties of GMM based Wald statistics within the confines of a very simple statistical environment. In particular we suppose that data generating process is a mean zero, unit variance Gaussian white noise process. There are several advantages to working with such a simple process. First, we are able to document that the basic problems which arise in the more complex environment considered in section 3 also arise here. Second, developing intuition for the results is easier in a simpler environment. Third, we can examine the effects of imposing various assumptions about the data generating processes on our procedures. Fourth, we can compute all relevant population moments exactly. Fifth, simulation is straightforward and the number of replications can be increased to gain accuracy in our Monte Carlo experiments.

The remainder of this section is organized as follows. Subsection 2.1 describes the data generating process. In subsections 2.2 and 2.3 we discuss the hypothesis tests and different experiments that we conducted. Finally, we report the results of our Monte Carlo experiments in subsection 2.4.

2.1 The Data Generating Process

We suppose that an econometrician has time series data on $J = 20$ random variables X_{it}, $i = 1, \ldots, J$, each of which are i.i.d. N(0,1) and mutually independent.[2] The econometri-

---

[2] We also conducted experiments in which the data were independent MA(1) processes with Gaussian innovations, and which were either positively or negatively serially correlated. In both cases our results were

cian has $T = 100$ observations on X_{it}, $i = 1, \ldots, J$. To simplify the analysis we assume that the econometrician knows that $EX_{it} = 0$, for all i and t. The econometrician is interested in estimating and testing hypotheses about the standard deviations, σ_i, of X_{it}, $i = 1, 2, \ldots, J$. To estimate σ_i he uses a simple exactly identified GMM estimator based on the moment restriction

$$E(X_{it}^2 - \sigma_i^2) = 0, \quad i = 1, 2, \ldots, J. \tag{1}$$

This leads to the GMM estimators

$$\hat{\sigma}_i = \left(\frac{1}{T} \sum_{t=1}^{T} X_{it}^2 \right)^{1/2}. \tag{2}$$

2.2 Hypothesis Testing

The econometrician estimates σ_i in order to conduct inference. The hypotheses of interest pertain to the variability of the series X_{it}. The specific hypotheses to be tested are of the form

$$H_M : \sigma_1 = \sigma_2 = \cdots = \sigma_M = 1, \quad M \leq J.$$

We consider this hypothesis because of its similarity to a diagnostic procedure that is often used to evaluate RBC (and other) models. The basic idea is to see whether a model can 'account' for various second moments of the data. In practice this amounts to testing whether the second moments of some series estimated in a nonparametric manner equal the analogous second moment implications of a particular RBC model (see section 3). Early work on RBC models tended to concentrate on the volatility of different economic aggregates (see for example Hansen (1985)). Here there is no 'model'. But we can test sample moments against their true value ($M = 1$) and test whether various second moments are equal to each other in population using similar statistical procedures.

The specific Wald statistic that we use to test H_M is given by

$$\mathcal{W}_T^M = T(\hat{\sigma} - 1)' A' (A V_T A')^{-1} A (\hat{\sigma} - 1). \tag{3}$$

Here $A = \left(\begin{array}{cc} I_M & 0_{M \times J - M} \end{array} \right)$ and V_T denotes a generic estimator of the asymptotic variance-covariance matrix of $\sqrt{T}(\hat{\sigma} - \sigma_0)$, where σ_0 is the true value of the parameter vector $\sigma = \left(\begin{array}{cccc} \sigma_1 & \sigma_2 & \cdots & \sigma_J \end{array} \right)'$. Given well behaved estimators $\hat{\sigma}$ and V_T, $\mathcal{W}_T^M \xrightarrow{d} \chi^2(M)$.

qualitatively similar to the white noise case.

4

We consider several questions that arise in testing H_M. First, how does the choice of estimator V_T affect inference? We are particularly interested in assessing the small sample implications of using non-parametric estimators of V_T and understanding the gains to imposing different types of restrictions on V_T. Particularly important sources of restrictions are the economic theory being investigated and the null hypothesis being tested. For example, intertemporal consumption based asset pricing models typically imply restrictions on the degree of serial correlation in the error terms that define V_T. (See for example Hansen and Singleton (1982) or Eichenbaum and Hansen (1990)). A different example is provided in section 3 where we can use the structural model itself to generate an estimate of V_T. Since imposing restrictions on V_T can often be computationally burdensome, and asymptotic inference is not affected, it is important to understand the nature of the small sample gains to doing so.

Second, how does the dimension of the test, i.e. the degrees of freedom M, affect the size of the test? This question is important because, in many applications, the model gives rise to a large number of over-identifying restrictions. The issue is what trade-offs are involved in simultaneously testing more or less of these moment restrictions.

Third, how are the small sample properties of the Wald statistic affected by reparameterizing the example?[3] An asymptotically equivalent way of assessing hypothesis H_M is to proceed as follows. Suppose that we estimate σ_1, along with $\theta_i = \sigma_i/\sigma_1$ for $i = 2, 3, \ldots, J$. To estimate θ_i we utilize the following moment restrictions.

$$
\begin{aligned}
E(X_{1t}^2 - \sigma_1^2) &= 0 \\
E(X_{it}^2 - \theta_i^2 X_{1t}^2) &= 0. \quad i = 2, \ldots, J
\end{aligned}
\tag{4}
$$

This leads to the estimators

$$
\hat{\sigma}_1 = \left(\frac{1}{T} \sum_{t=1}^{T} X_{1t}^2 \right)^{1/2}
$$

$$
\hat{\theta}_i = \left(\frac{1}{T} \sum_{t=1}^{T} X_{it}^2 \Big/ \frac{1}{T} \sum_{t=1}^{T} X_{1t}^2 \right)^{1/2}. \quad i = 2, \ldots, J
$$

The analogous hypothesis to H_M is

$$
\tilde{H}_M : \sigma_1 = \theta_2 = \cdots = \theta_M = 1, \quad M \leq J.
$$

[3]Gregory and Veall (1985) study the effects of reparameterizing Wald tests in a regression context.

5

The corresponding Wald test statistic for this hypothesis is

$$\widetilde{\mathcal{W}}_T^M = T(\hat{\theta} - 1)'A'(A\bar{V}_T A')^{-1}A(\hat{\theta} - 1), \tag{5}$$

where $\theta = \begin{pmatrix} \sigma_1 & \theta_2 & \cdots & \theta_J \end{pmatrix}'$ and $\bar{V}_T$ is some estimator for the asymptotic variance-covariance matrix of $\hat{\theta}$. There is no a priori reason to suppose that the small sample properties of $\mathcal{W}_T^M$ will be the same as those of $\widetilde{\mathcal{W}}_T^M$.

This example is of interest because it can shed light on the common practice in the RBC literature of testing whether a model matches the volatility of output and the volatility of various aggregates *relative* to output. One could simply test whether a model matches the absolute volatility of all the relevant variables. Asymptotically this choice should not matter. But the small sample properties of the Wald tests in the two cases could be quite different.

2.3 Alternative Covariance Matrix Estimators

In this section we discuss our estimators of the asymptotic variance-covariance matrix of $\hat{\sigma}$ and $\hat{\theta}$. To be concrete we concentrate on the case of $\hat{\sigma}$. The case of $\hat{\theta}$ is discussed in Appendix A. The moment conditions used to estimate σ, (1), can be written in the form $E[g(X_t, \sigma)] = 0$. Here $g(\cdot, \cdot)$ is the $J \times 1$ vector valued function whose ith element is given by $(X_{it}^2 - \sigma_i^2)$. Denoting the true value of σ by σ_0, the asymptotic covariance matrix of $\sqrt{T}(\hat{\sigma} - \sigma_0)$ is given by

$$V_0 = (D_0'S_0^{-1}D_0)^{-1},$$

where

$$D_0 = E\frac{\partial g(X_t, \sigma_0)}{\partial \sigma'},$$

and

$$S_0 = \sum_{j=-\infty}^{\infty} g(X_t, \sigma_0)g(X_{t-j}, \sigma_0)'.$$

The corresponding estimator of V_0 is given by

$$V_T = (D_T'S_T^{-1}D_T)^{-1},$$

where D_T and S_T are consistent estimators for D_0 and S_0.

6

We consider several estimators of V_0. Each is defined in terms of some estimators D_T and S_T. The different estimators impose varying amounts of information at the econometrician's disposal. Some of this information is in the nature of the maintained assumptions concerning the serial and mutual independence properties of X_{it} and Gaussianity. Other information derives from the null hypothesis being tested.

Initially we consider estimators of S_0 which do not exploit any of this information. Instead we estimate S_0 using versions of the nonparametric estimator proposed by Newey and West (1987).[4] A general version of this estimator can be written as

$$S_T = \sum_{j=-(T-1)}^{T-1} k(\frac{j}{B_T})\hat{\Omega}_j,$$

where

$$\hat{\Omega}_j = \begin{cases} (1/T)\sum_{t=j+1}^{T} g(X_t, \hat{\sigma})g(X_{t-j}, \hat{\sigma})' & \text{for } j \geq 0 \\ (1/T)\sum_{t=-j+1}^{T} g(X_{t+j}, \hat{\sigma})g(X_t, \hat{\sigma})' & \text{for } j < 0 \end{cases}$$

and

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

Here B_T is a scalar that determines the bandwidth of the lag window $k(\cdot)$. We consider three variants of this estimator,

- S_T^1 uses bandwidth $B_T = 4$,

- S_T^2 uses $B_T = 2$,

- S_T^3 has B_T chosen automatically using a procedure suggested by Andrews (1991) which is described in more detail in Appendix C.

The next group of estimators that we consider utilizes additional amounts of information about the underlying data generating process. The estimator S_T^4 exploits the assumption that the X_{it} are serially uncorrelated. This implies that

- S_T^4 has ijth element given by $\frac{1}{T}[\sum_{t=1}^{T}(X_{it}^2 - \hat{\sigma}_i^2)(X_{jt}^2 - \hat{\sigma}_j^2)]$.

The estimator S_T^5 imposes the mutual independence of the X_{it}'s as well as their serial independence. This implies that

---

[4]There are several alternative estimators which could be used at this stage. We have found our results to be relatively insensitive to the choice of procedure in the RBC context so we present results based only on the Newey and West (1987) method.

7

- S_T^5 is a diagonal matrix with iith element given by $\frac{1}{T}[\sum_{t=1}^{T}(X_{it}^2 - \hat{\sigma}_i^2)^2]$.

Our next estimator, S_T^6, also exploits the fact that the X_{it} are Gaussian. Since Gaussianity implies that $E(X_{it}^4) = 3\sigma_i^4$,

- S_T^6 is a diagonal matrix with iith element given by $2\hat{\sigma}_i^4$.

Our next two estimators impose additional restrictions derived from the null hypothesis being tested. Under hypothesis H_M, $\sigma_i = 1$ for $i = 1, \ldots, M$, while σ_i is unrestricted for $i = M + 1, \ldots, J$. This suggests the estimator

- S_T^7 which is a diagonal matrix with iith element 2 for $i \leq M$, and $2\hat{\sigma}_i^4$ for $i = M + 1, \ldots, J$.

Corresponding to each estimator of S_0 discussed above, there is an estimator for V_0 given by,

$$V_T^k = (D_T^{1\prime}(S_T^k)^{-1}D_T^1)^{-1},$$

$k = 1, 2, \ldots, 7$, where D_T^1 is a diagonal matrix with iith element $-2\hat{\sigma}_i$. Since the null hypothesis can also be imposed on D_T^1, we also consider the estimator

$$V_T^8 = (D_T^{2\prime}(S_T^7)^{-1}D_T^2)^{-1},$$

where D_T^2 is a diagonal matrix with iith element -2 for $i \leq M$, and $-2\hat{\sigma}_i$ for $i = M + 1, \ldots, J$.[5] Here the $\mathcal{W}$ statistic reduces to $\sum_{i=1}^{M}(\hat{\sigma}_i - 1)^2/2$.

We use the same differential information assumptions to define eight estimators for the variance-covariance matrix of $\hat{\theta}$ that are analogous to V_T^k, $k = 1, 2, \ldots, 8$ (see Appendix A).

2.4 Monte Carlo Experiments

Our experiments were conducted as follows. We generated 10,000 sets of synthetic time series on $\{X_{1t}, X_{2t}, \ldots, X_{Jt}\}_{t=1}^{T}$, each of length 100. On each artificial data set, we estimated the parameter vector σ, the different estimators of the variance covariance matrix

---

[5] If we imposed $\sigma_i = 1$ for all i in the computation of $\hat{V}_T^7$ and $\hat{V}_T^8$ we would get numerically identical results for our test statistics because all the matrices involved in the calculation are diagonal.

8

and then calculated the Wald test statistic, $\mathcal{W}_T^M$, that is relevant for testing hypothesis H_M, $M \in \{1, 2, 5, 10, 20\}$. This allowed us to generate an empirical distribution function for $\mathcal{W}_T^M$ under the null hypothesis that H_M is true, corresponding to the different estimators of V_0.

Our results are summarized in Table 1, the columns of which correspond to different specifications of M (which also equals the degrees of freedom of the test). The rows correspond to fixed asymptotic sizes of the test while the entries in the table are the percentages, out of 10,000 draws, for which the $\mathcal{W}$ statistic exceeded the relevant critical value of the chi-squared distribution.

A number of interesting results emerge here. Consider first the distribution of the test statistics generated using V_T^1, V_T^2, V_T^3 and V_T^4 (see Panels A–D of Table 1). First, even for $M = 1$, the small sample size of the tests exceeds their asymptotic size. This result is similar to that obtained by Christiano and den Haan (1994) and Newey and West (1993). Second, the small sample size of the tests rises uniformly with M. Indeed when we use the estimator V_T^1, the $\mathcal{W}$ statistic for hypothesis H_{20} exceeds its asymptotic (1%, 5%, 10%) critical values (59%, 73%, 80%) of the time. For even moderate sizes of M, relying on asymptotic distribution theory leads one to reject H_M far more often than is warranted in small samples. It is true that as the bandwidth decreases, the small sample performance of the Wald test improves uniformly. But as panel D indicates, even when we impose the white noise assumption (i.e. we use V_T^4), the small sample performance of the large joint tests is dismal. For example, with $M = 20$, tests with asymptotic size (1% , 5%, 10%), lead to rejection (17%, 33%, 43%) of the time in samples of 100 observations.

The results generated using V_T^5 (which exploits the assumption that the X_{it} are mutually independent) are presented in Panel E of Table 1. Comparing Panel E to Panels A–D, we see that the impact of imposing the independence assumption is to move the small sample sizes of the tests substantially closer to their asymptotic values. Not surprisingly, the impact of this restriction becomes larger as M increases since there are more off-diagonal elements being set to their population values. (In the case of $M = 1$ the two panels are identical). With $M = 20$, the $\mathcal{W}$ statistic for H_M exceeds its asymptotic (1%, 5%, 10%) critical values (4.7%, 13.4%, 21.2%) of the time. This represents a substantial improvement relative to the situation when we do not impose the zero off-diagonal element restriction. Even so, the Wald test still rejects too often in small samples. Panel F,

which reports results based on V_T^6, indicates that imposing the Gaussianity assumption improves the small performance of $\mathcal{W}_T^M$ even further. To the extent that fourth moments are less accurately estimated than second moments for Gaussian processes this result is not surprising.

Recall that the estimator V_T^7 exploits information from the null hypothesis regarding σ_i in constructing S_T^7. The results generated using V_T^7 are reported in Panel G of Table 1. Comparing Panels F and G we see that the net effect of imposing these additional restrictions is to move the small sample size of the test even closer to its asymptotic size (except for the 10% critical value for $M = 1$). For example, with $M = 20$, the $\mathcal{W}$ statistic for H_M exceeds its asymptotic (1%, 5%, 10%) critical values (2.1%, 7.3%, 12.1%) of the time.

Panel H of Table 1 reports results based on V_T^8 where we impose the null hypothesis on D_T as well as on S_T. Now all of the anomalies associated with the small sample distribution of the $\mathcal{W}$ statistic disappear. First, the degree to which the small sample sizes match their asymptotic sizes is not affected by M. Second, the small sample size of the test statistic is extremely close to the corresponding asymptotic size. Indeed, this is true even when we fix the asymptotic size of the test at 1%. So, at least for the present example, the parameter estimates appear to have a small sample distribution which is very well approximated by their large sample distribution. The problem with the small sample distribution of the $\mathcal{W}$ statistic seems to be closely related to the small sample distribution of S_T and to a much smaller extent D_T. The more information the econometrician imposes on S_T and D_T, the better the performance of the tests appears to be in this example.

Table 2 presents results pertaining to the $\widetilde{\mathcal{W}}_T$ statistic that is relevant for our alternative parameterization of the problem in terms of relative standard deviations.[6] In many ways these results are qualitatively similar to those obtained with the original parameterization. Broadly speaking, the second set of tests leads to slightly more rejections, although only to a modest extent. However, unlike the previous parameterization, when we impose all of the available information on S_T and D_T (Panel H of Table 2), there is still a noticeable tendency of joint tests with many degrees of freedom to reject more frequently than tests of single hypotheses.

---

[6] The column in Table 2 headed $M = 1a$ is for tests of the hypothesis $\sigma_1 = 1$, while the column headed $M = 1b$ is for tests of the hypothesis $\sigma_2/\sigma_1 = 1$. Both tests have one degree of freedom.

These results suggest that simply reparameterizing the problem will not dramatically improve the performance of tests constructed using nonparametric estimators for S_T and D_T. The key problem with inference seems to arise from difficulties in estimating the spectral density matrix of the GMM error terms, S_0. Imposing as much information as possible when estimating S_0 and D_0 leads to significant improvements in the size properties of the Wald tests. In the next section we investigate the extent to which these conclusions continue to hold in a more complex statistical environment.

3 A Real Business Cycle Model As a Data Generating Process

In this section we consider the small sample properties of GMM based Wald statistics assuming that the data generating process is given by the business cycle model developed in Burnside and Eichenbaum (1994). The model is briefly summarized in subsection 3.1. Subsection 3.2 describes the way the model's structural parameters were estimated. Subsection 3.3 discusses the hypothesis tests we investigated. In subsection 3.4 we present the results of our Monte Carlo experiments.

3.1 The Model

The model economy is populated by a large number of infinitely lived individuals. To go to work an individual must incur a fixed cost of ς hours. Once at work, an individual stays for a fixed shift length of f hours. The time t instantaneous utility of such a person is given by

$$\ln(C_t) + \theta \ln(T - \varsigma - W_t f) \tag{6}$$

Here T denotes the individual's time endowment, C_t denotes time t privately purchased consumption, $\theta > 0$, and W_t denotes the time t level of effort. The time t instantaneous utility of a person who does not go to work is given by $\ln(C_t) + \theta \ln(T)$.

Time t output, Y_t, is produced via the Cobb-Douglas production function

$$Y_t = (K_t U_t)^{1-\alpha} (N_t f W_t X_t)^\alpha \tag{7}$$

where $0 < \alpha < 1$, K_t denotes the beginning of time t capital stock, U_t represents the capital utilization rate, N_t denotes the number of individuals at work during time t, and

11

X_t represents the time t level of technology. We assume that the time t depreciation rate of capital, δ_t, is given by

$$\delta_t = \delta U_t^\phi \tag{8}$$

where $0 < \delta < 1$ and $\phi > 1$. The stock of capital evolves according to

$$K_{t+1} = (1 - \delta_t)K_t + I_t \tag{9}$$

where I_t denotes time t gross investment.

The level of technology, X_t, evolves according to

$$X_t = X_{t-1}\exp(\gamma + v_t)$$

where v_t is a serially uncorrelated process with mean 0 and standard deviation σ_v. The aggregate resource constraint is given by

$$C_t + I_t + G_t \le Y_t \tag{10}$$

where G_t denotes the time t level of government consumption. We assume that G_t evolves according to

$$G_t = X_t g_t^* \tag{11}$$

Here g_t^* is the stationary component of government consumption and $g_t = \ln(g_t^*)$ evolves according to

$$g_t = \mu + \rho g_{t-1} + \epsilon_t \tag{12}$$

where μ is a scalar, $|\rho| < 1$ and ϵ_t is a serially uncorrelated process with mean 0 and standard deviation σ_ϵ.

In the presence of complete markets the competitive equilibrium of this economy corresponds to the solution of the social planning problem:

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[\ln(C_t) + \theta N_t \ln(T - \xi - W_t f) + \theta(1 - N_t)\ln(T)\right] \tag{13}$$

subject to (7) - (12) by choice of contingency plans for $\{C_t, K_{t+1}, N_t, U_t, W_t : t \ge 0\}$. We obtain an approximate solution to this problem using King, Plosser and Rebelo's (1988) log-linear solution procedure.[7] Let $k_t = \ln(K_t/X_{t-1})$, $h_t = \ln(H_t)$, $c_t = \ln(C_t/X_t)$,

---

[7]See Burnside (1993) for details.

$w_t = \ln(W_t)$, $u_t = \ln(U_t)$, $y_t = \ln(Y_t/X_t)$, $a_t = \ln(Y_t/N_t X_t)$, $i_t = \ln(I_t/X_t)$, $h_t^o = \ln(H_t^o)$, and $a_t^o = \ln(Y_t/H_t^o X_t)$. Here H_t and H_t^0 denote actual and observed time t hours of work.

As in Prescott (1986), we assume that

$$\ln(H_t^0) = \ln(H_t) + \xi_t \tag{14}$$

where ξ_t is an i.i.d. random variable with mean zero and variance σ_ξ^2. The time t state of the system is given by

$$s_t = \begin{pmatrix} 1 & k_t & h_t & v_t & g_t & \xi_t \end{pmatrix}'$$

Define the vector of time t endogenous variables f_t as

$$f_t = \begin{pmatrix} c_t & w_t & u_t & y_t & a_t & i_t & h_t^o & a_t^o \end{pmatrix}'$$

and the vector of time t shocks

$$\tilde{\epsilon}_t = \begin{pmatrix} 0 & 0 & 0 & v_t & \epsilon_t & \xi_t \end{pmatrix}'.$$

Our assumptions about the exogenous variables and the log-linear approximation to the model imply that the evolution of the system can be summarized as

$$\begin{aligned} s_t &= M s_{t-1} + \tilde{\epsilon}_t \\ f_t &= \Pi s_t \end{aligned} \tag{15}$$

where M and Π are functions of the model's underlying structural parameters. We take (15) to be the data generating mechanism in our Monte Carlo experiments.[6]

3.2 Estimation

With certain exceptions, the parameters of the model were estimated using a variant of the GMM procedure described in Christiano and Eichenbaum (1992). We did not estimate β, T, ς and f. Instead we set $\beta = 1.03^{-1/4}$, $T = 1369$ hours per quarter, $\varsigma = 60$ and chose f so that the nonstochastic steady state value of W_t is 1. Rather than estimating the parameter δ, we estimated $\tilde{\delta} = \delta U^\phi$, where U is the nonstochastic steady state value of U_t. To obtain a value of ϕ we use the fact that in nonstochastic steady state,

$$\phi = \frac{\beta^{-1}\exp(\gamma) - 1}{\tilde{\delta}} + 1.$$

13

In the data, the series g_t displays a time trend, so this series was detrended using a linear time trend. To simplify matters we did not include the time trend in the Monte Carlo experiments. In addition, we chose to estimate the nonstochastic steady state value of G_t/Y_t, as the parameter g/y, rather than the mean of the process $g_t = \ln(G_t/X_t)$.[9]

In light of these decisions, the vector of model parameters to be estimated, denoted by Ψ_1, is given by

$$\Psi_1 = \left(\begin{array}{ccccccccc} \theta & \alpha & \delta & \gamma & \sigma_v & g/y & \rho & \sigma_\epsilon & \sigma_\xi \end{array} \right)'.$$

The hypotheses that we investigate involve various second moments of the data. Since many of the relevant series exhibit marked trends, some stationary inducing transformation of the data must be applied. To facilitate comparisons with the RBC literature, we chose to process the data using the Hodrick and Prescott (1980) filter. Consequently, the second moments to be discussed pertain to those of Hodrick and Prescott (HP) filtered data.[10]

We focus on a set of second moments that have received a great deal of attention in the RBC literature: the standard deviation of output, σ_y, the standard deviations of consumption, investment and hours relative to the standard deviation of output, σ_c/σ_y, σ_i/σ_y and σ_h/σ_y and the standard deviation of hours worked relative to the standard deviation of average productivity, σ_h/σ_a. We also consider the dynamic correlations between average productivity and hours, $\rho_{ah}^i = \text{Corr}(APL_t, H_{t+i})$, $i = \pm 1, \pm 2, \pm 3, \pm 4$, and the dynamic correlations between average productivity and output $\rho_{ay}^i = \text{Corr}(APL_t, Y_{t+i})$, $i = -4, -3, -2, -1$.[11] We denote the vector of diagnostic moments that must be estimated in ways not involving the model by

$$\Psi_2 = \left(\begin{array}{cccccccccc} \sigma_y & \sigma_c/\sigma_y & \sigma_i/\sigma_y & \sigma_h/\sigma_y & \sigma_h/\sigma_a & \rho_{ah}^{-4} & \rho_{ah}^{-3} & \rho_{ah}^{-2} & \rho_{ah}^{-1} \\ & & \rho_{ah}^{1} & \rho_{ah}^{2} & \rho_{ah}^{3} & \rho_{ah}^{4} & \rho_{ay}^{-4} & \rho_{ay}^{-3} & \rho_{ay}^{-2} & \rho_{ay}^{-1} \end{array} \right)'.$$

---

[9] The mean of g_t would matter in the linearised solution only in determining the steady state share of government expenditure in output, which we parameterise directly.

[10] We have redone all of the experiments in this paper with first differenced data. For a comparison of some of the small sample properties of GMM with HP-filtered and first differenced data see Christiano and den Haan (1994).

[11] The contemporaneous correlations between these variables and ρ_{ay}^i, $i = 1, 2, 3, 4$, can be deduced from the other moments that we consider.

3.2.1 Moment Conditions Underlying the Estimator of Ψ_1

As discussed in Burnside and Eichenbaum (1994), our estimator of Ψ_1 is based on the following moment conditions:

$$E\left\{\sigma_\xi^2 - \frac{1}{2}\left[\Delta\ln(H_t^o)\right]^2 + \frac{1}{2}\Delta\ln(H_t^o)\Delta\ln(\tilde{H}_t^o)\right\} = 0 \tag{16}$$

$$E\left[\ln(H_t^o) - \ln(Nf)\right] = 0 \tag{17}$$

$$E\left[\ln(\delta_t) - \ln(\delta)\right] = 0 \tag{18}$$

$$E\left[1 - \beta\left(\frac{C_t}{C_{t+1}}\right)\left[(1-\alpha)(1-\phi^{-1})\frac{Y_{t+1}}{K_{t+1}} + 1\right]\right] = 0 \tag{19}$$

$$E\left[\ln(X_t^o) - \ln(X_{t-1}^o) - \gamma\right] = 0 \tag{20}$$

$$E\left[\ln(X_t^o) - \ln(X_{t-1}^o)\right]^2 - \sigma_s^2 - 2\varphi_3^2\sigma_\xi^2 = 0 \tag{21}$$

$$E\left[\ln(G_t) - \ln(Y_t) - \ln(g/y)\right] = 0 \tag{22}$$

$$E\left[(g_t^o - \rho g_{t-1}^o)\ln(g_{t-1}^o)\right] + \rho\varphi_3^2\sigma_\xi^2 = 0 \tag{23}$$

$$E\left[(g_t^o - \rho g_{t-1}^o)^2\right] - (1+\rho^2)\varphi_3^2\sigma_\xi^2 - \sigma_t^2 = 0 \tag{24}$$

In equation (16), H_t^0 and $\tilde{H}_t^o$ refer to our two measures of hours worked (see Appendix D).[12] The variables N, representing the nonstochastic steady state value of N_t, and φ_3, a reduced form parameter, are functions of the underlying parameter vector, Ψ_1. Furthermore, X_t^o represents a measurement-error corrupted signal of the level of technology which can be constructed given the data and a vector of parameters Ψ_1. Similarly, g_t^o is a signal of g_t based on the error-ridden measure of technology X_t^o.[13]

3.2.2 Moment Conditions Underlying the Estimator of Ψ_2

Our estimator of Ψ_2 is based on the following moment conditions:

$$E\left[y_t^2 - \sigma_y^2\right] = 0$$

$$E\left[c_t^2 - (\sigma_c/\sigma_y)^2 y_t^2\right] = 0$$

$$E\left[i_t^2 - (\sigma_i/\sigma_y)^2 y_t^2\right] = 0$$

---

[12] Unlike Burnside and Eichenbaum (1994), we abstract from issues concerning the observability of δ_t and K_t. In particular, we assume, for the purposes of our Monte Carlo experiments, that the econometrician observes these series directly.

[13] See Burnside and Eichenbaum (1994) for details.

$$E\left[h_t^2 - (\sigma_h/\sigma_y)^2 y_t^2\right] = 0$$

$$E\left[h_t^2 - (\sigma_h/\sigma_a)^2 a_t^2\right] = 0$$

$$E\left[a_t h_{t+i} - \rho_{a,h}^i \left(\frac{\sigma_h}{\sigma_y}\right)^2 \sigma_y^2/\left(\frac{\sigma_h}{\sigma_a}\right)\right] = 0, \; i = \pm 1, \ldots, \pm 4$$

$$E\left[a_t y_{t+i} - \rho_{a,y}^i \left(\frac{\sigma_h}{\sigma_y}\right) \sigma_y^2/\left(\frac{\sigma_h}{\sigma_a}\right)\right] = 0, \; i = 1, \ldots, 4,$$

where a lower case variable, e.g. z_t, is the cyclical component of $\ln(Z_t)$ as defined by the HP filter.

To define our joint estimator of Ψ_1 and Ψ_2 consider the following generic representation of our moment conditions:

$$E\left[u_t(\Psi^0)\right] = 0 \qquad t = 1, \ldots, T$$

where Ψ^0 is the true value of $\left(\begin{array}{cc} \Psi_1' & \Psi_2' \end{array}\right)'$ and u_t is a vector valued function of dimension equal to the dimension of Ψ^0. Let

$$g_T(\Psi) = \frac{1}{T} \sum_{t=1}^{T} u_t(\Psi).$$

The GMM estimator, Ψ_T, minimizes

$$J_T = T g_T(\Psi)' \Upsilon_T g_T(\Psi). \tag{25}$$

where Υ_T is a symmetric positive definite weighting matrix of dimension equal to the dimension of $g_T(\Psi)$. Since our GMM estimator is exactly identified, Ψ_T is independent of Υ_T. We simply set Υ_T equal to the identity matrix in (25).

A consistent estimator of the variance-covariance matrix of $\sqrt{T}(\Psi_T - \Psi_0)$ is given by

$$V_\Psi = \left(D_T' S_T^{-1} D_T\right)^{-1}$$

where $D_T = \partial g_T(\Psi_T)/\partial \Psi'$ and S_T is a consistent estimate of S_0, 2π times the spectral density matrix of $u_t(\Psi^0)$ at frequency zero.

3.3 Hypothesis Testing

Suppose we wish to assess the empirical plausibility of the model's implications for a $q \times 1$ subset of Ψ_2 given by ω. Let $\Phi(\Psi)$ denote the value of ω implied by the model, given the

16

structural parameters Ψ_1. Here Φ denotes the (nonlinear) mapping between the model's structural parameters and the relevant population moments. Denote the nonparametric estimate of ω obtained without imposing restrictions from the model by $\Gamma(\Psi)$.[14] The hypotheses that we investigate are of the form

$$H_0 : F(\Psi^0) = \Phi(\Psi^0) - \Gamma(\Psi^0) = 0 \tag{26}$$

Christiano and Eichenbaum (1992) show that a consistent estimate of the asymptotic variance-covariance matrix of $\sqrt{T}[F(\Psi_T) - F(\Psi_0)]$ is

$$V_F = \left[\frac{\partial F(\Psi_T)}{\partial \Psi'}\right] V_\bullet \left[\frac{\partial F(\Psi_T)}{\partial \Psi'}\right]'$$

and that the test statistic

$$\mathcal{W}_T = T F(\Psi_T)' V_F^{-1} F(\Psi_T) \tag{27}$$

is asymptotically distributed as a χ^2 random variable with q degrees of freedom.

We consider two types of hypothesis tests. The first type involve tests of individual moments of the data. The test numbers and corresponding moments being tested are summarized in the following table.

Test #	Moment	Test #	Moment	Test #	Moment
1	σ_y	6	ρ_{ah}^{-4}	15	ρ_{ay}^{-4}
2	σ_c/σ_y	7	ρ_{ah}^{-3}	16	ρ_{ay}^{-3}
3	σ_i/σ_y	8	ρ_{ah}^{-2}	17	ρ_{ay}^{-2}
4	σ_h/σ_y	9	ρ_{ah}^{-1}	18	ρ_{ay}^{-1}
5	σ_h/σ_a	10	ρ_{ah}^{0}	19	ρ_{ay}^{0}
		11	ρ_{ah}^{1}	20	ρ_{ay}^{1}
		12	ρ_{ah}^{2}	21	ρ_{ay}^{2}
		13	ρ_{ah}^{3}	22	ρ_{ay}^{3}
		14	ρ_{ah}^{4}	23	ρ_{ay}^{4}

The second type of tests involve joint moment restrictions. Hypothesis H1 states that the values of σ_y, σ_c/σ_y, σ_i/σ_y, σ_h/σ_y and σ_h/σ_a implied by the model are the same, in population, as the corresponding moments of the data generating process. Hypotheses H2, H3 and H4 are similar to hypothesis H1 but pertain to the moments $\{\rho_{ah}^i, \ i =$

---

[14] Often the mapping Γ is linear. In particular, Γ is often a conformable matrix of zeros and ones that selects the vector ω from Ψ_2.

$0, \pm 1, \pm 2, \pm 3, \pm 4\}$, $\{\rho^i_{\alpha y}, \ i = 0, \pm 1, \pm 2, \pm 3, \pm 4\}$ and $\{\sigma_y, \ \sigma_c/\sigma_y, \ \sigma_i/\sigma_y, \ \sigma_h/\sigma_y, \ \sigma_h/\sigma_a, \rho^i_{\alpha h},$
$i = \pm 1, \pm 2, \pm 3, \pm 4, \ \rho^i_{\alpha y}, \ i = -4, -3, -2, -1\}$, respectively.[15] The test statistics for H1,
H2, H3 and H4 have 5, 9, 9, and 17 degrees of freedom, respectively.

To implement our hypotheses tests we require an estimator, S_T, of S_0. As in section
2, our estimators are of the form

$$S_T = \sum_{j=-(T-1)}^{T-1} k\left(\frac{j}{B_T}\right) \hat{\Omega}_j,$$

where

$$\hat{\Omega}_j = \begin{cases} (1/T) \sum_{t=j+1}^{T} \hat{u}_t \hat{u}'_{t-j} & \text{for } j \geq 0 \\ (1/T) \sum_{t=-j+1}^{T} \hat{u}_{t+j} \hat{u}'_t & \text{for } j < 0 \end{cases}$$

The kernel function k varies depending on the estimator, B_T is the bandwidth and $\hat{u}_t = u_t(\Psi_T)$. Our baseline results are generated using the Bartlett kernel function

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

and Andrews' (1991) automatic selection procedure for B_T. Appendix C discusses the
other estimators of S_0 that we consider. As it turns out, our basic results are robust
across these different estimators of S_0.

The bandwidth selection procedure that we used can be described as follows. Andrews
(1991) provides an expression for the optimal bandwidth corresponding to a given kernel,
a process u_t, and a set of weights on the different elements of S_0. The bandwidth is
optimal in the sense that it leads to minimum MSE estimates of a weighted inner product
of the elements of S_0. Andrews' (1991) procedure simplifies the dependence of the optimal
bandwidth on the entire spectral density of u_t by assuming a simple parametric model for
the error term. The choice of model does not affect the consistency of S_T. The model
which we use corresponds to the simplest example in Andrews (1991). Specifically, we
treat the elements of u_t as independent AR(1) scalar processes. No weight is given to
the off-diagonal elements of S_0. Under these circumstances, the bandwidth selected will
depend on the sample size, T, the weights, and coefficient estimates obtained by fitting
AR(1) processes to the elements of $u_t(\Psi_T)$. Roughly speaking, the more persistent the
errors, the greater the bandwidth.

---

[15]The last set of moments contains the nonredundant elements from among the moments involved in tests
1-23.

In the standard case, equal weight is placed on all of the error terms. However we found that doing this led to test statistics with very poor small sample properties. (See Appendix C). Instead we placed zero weight on (17), (18) and (22) along with unit weight on the other error terms. The resulting median bandwidth across the different Monte Carlo draws was 2.78.

3.4 Parameter Estimates and Some Results Based on Asymptotic Theory

Table 3 reports our point estimates of Ψ_1 along with corresponding standard errors. The data set used to generate these estimates is described in Appendix D. Table 4 presents the non-model and model based estimates of $\{\sigma_y, \sigma_c/\sigma_y, \sigma_i/\sigma_y, \sigma_h/\sigma_y, \sigma_h/\sigma_a\}$. Numbers in parentheses are the standard errors of the corresponding point estimates. Numbers in brackets are the asymptotic probability values of the $\mathcal{W}$ statistics for testing whether the individual model and data population moment are the same. Notice that we cannot reject any of the individual hypotheses in question.

Figures 1 and 2 summarize the model's implications for the dynamic correlations between hours worked and average productivity as well as the dynamic correlations between average productivity and output. The dotted lines in row 1 correspond to the non-model based estimates of $\{\rho_{aA}^i, i = 0, \pm 1, \pm 2, \pm 3, \pm 4\}$, and $\{\rho_{ay}^i, i = 0, \pm 1, \pm 2, \pm 3, \pm 4\}$, while the solid lines denote the moments implied by the model. The solid lines in row 2 graph the differences between the model and non-model based estimates while the dotted lines depict an asymptotic two standard error band for the differences. According to these figures, the model does quite well at accounting for the individual dynamic correlations between average productivity and output as well as average productivity and hours worked.

We now turn to our joint hypotheses. Columns 1 and 2 of Table 5 report the $\mathcal{W}$ statistics for hypotheses {H1, H2, H3, H4} and the corresponding asymptotic probability values. Notice that hypotheses H2, H3 and H4 are all rejected at very low significance levels. To us the strength of these rejections seems at variance with the results of testing the individual components of these hypotheses. One way to reconcile these results is to invoke the pattern of covariances in question. However, in light of the results in section 2, these strong rejections may simply reflect the small sample properties of GMM based Wald statistics as applied to hypotheses involving joint moment restrictions. With this as

19

motivation we turn to the Monte Carlo experiments

3.5 Monte Carlo Experiments

To generate data for our Monte Carlo experiments we proceeded as follows. Given the estimated value of Ψ_1, we generated artificial time series according to the following rules: $C_t = \exp(c_t)X_t$, $Y_t = \exp(y_t)X_t$, $K_t = \exp(k_t)X_{t-1}$, $G_t = \exp(g_t)X_t$, $I_t = \exp(i_t)X_t$, $H_t = f\exp(n_t)$ and $\delta_t = \delta\exp(\phi u_t)$. Here c_t, y_t, k_t, u_t, n_t and i_t are given by (15). The variables X_t and g_t were generated according to the laws of motion specified in section 3.1. One thousand artificial time series data sets, each of length 113, were generated, assuming that the stochastic elements of $\tilde{\epsilon}_t$ were normally distributed.[16]

We begin by reporting the small sample behavior of the $\mathcal{W}$ statistics for hypotheses H1, H2, H3 and H4. Column 3 of Table 5 reports the percentage of times (out of 1000 Monte Carlo trials) that the $\mathcal{W}$ statistics for these hypotheses were greater than or equal to the corresponding $\mathcal{W}$ statistic obtained using U.S. data (see column 1). We refer to this fraction as the Monte Carlo probability. For hypothesis H1, H2 and H3 the asymptotic and Monte Carlo probabilities are reasonably similar. However for hypothesis H4 the Monte Carlo probability is much larger than the asymptotic probability (.06 versus .00). According to asymptotic standard distribution theory, the $\mathcal{W}$ statistic which we obtained for hypothesis H4 would be very unlikely if the model were specified correctly. But according to the small sample results, one would obtain a $\mathcal{W}$ statistic this large or larger roughly 6% of the time.

A complementary way to assess the small sample properties of the Wald tests is to consider the fraction of the time that the $\mathcal{W}$ statistics emerging from the Monte Carlos exceed the 1%, 5% and 10% critical values of the relevant chi-squared distributions. These are displayed in columns 4, 5 and 6 of Table 5. Notice that the small sample sizes of the test statistics for hypotheses H1 and H4 greatly exceed their asymptotic size. This tendency is particularly dramatic in the case of H4, where the $\mathcal{W}$ statistics exceed their asymptotic 1%, 5% and 10% critical values 37%, 51% and 58% of the time.

---

[16] With one exception all the moment conditions underlying our estimator of Ψ_1 hold exactly for the artificial data generating process. The exception is the planner's Euler equation for K_{t+1}, equation (19), discussed in Appendix B. To deal with this problem, we computed the expectation in equation (19) for the true log-linearized model. As it turns out, at these parameter values the error is approximately equal to 2×10^{-5}. To correct for possible bias we implemented our Monte Carlos centering equation (19) around 2×10^{-5} rather than 0.

Before analyzing this finding, we briefly discuss the size of the test statistics applied to the individual moments that make up joint hypotheses H1, H2, H3 and H4. Our results are displayed in Figure 3. The height of each bar graph in Panels A, B and C denotes the percentage of times (out of 1000 trials) that the $\mathcal{W}$ statistic for a given hypothesis exceeded the 10%, 5% and 1% critical values, of the asymptotic chi-squared distribution.

According to Figure 3, the small sample sizes of the test statistics for hypotheses 1 and 4-25 are moderately higher than their asymptotic sizes. The small sample sizes of the test statistics associated with σ_c/σ_y and σ_i/σ_y are substantially larger than their asymptotic sizes. This is consistent with our finding that Wald tests of hypotheses H1 and H4 over-reject in small samples. However, these effects do not seem large enough to explain the *extent* to which the Wald test over-rejects H1 and H4.

Viewed overall, the outstanding feature of our experiments is the large (small sample) size of the Wald test of hypothesis H4. Inference based on the asymptotic distribution of the $\mathcal{W}$ statistic leads to a grossly overly critical assessment of the model's performance. In Appendix C we show that this conclusion is robust to various perturbations. First, we consider the effects of different bandwidths when constructing S_T. These were chosen both on an *a priori* basis and using the Newey and West (1993) automatic bandwidth procedure. Second, we consider different estimators of S_0 that correspond to different lag windows. Third, we discuss the impact of using a small sample correction suggested by Andrews (1991).

A different dimension along which our results could be sensitive is how we parameterize the elements of Ψ_2. Specifically, we could include the moments $\{\sigma_c, \sigma_i, \sigma_h \sigma_a\}$ rather than $\{\sigma_c/\sigma_y, \sigma_i/\sigma_y, \sigma_h/\sigma_y, \sigma_h/\sigma_a\}$. Under these circumstances the moment conditions defining our estimator of Ψ_2 are given by:

$$E\left[y_t^2 - \sigma_y^2\right] = 0$$
$$E\left[c_t^2 - \sigma_c^2\right] = 0$$
$$E\left[i_t^2 - \sigma_i^2\right] = 0$$
$$E\left[h_t^2 - \sigma_h^2\right] = 0$$
$$E\left[a_t^2 - \sigma_a^2\right] = 0$$
$$E\left[a_t h_{t+i} - \rho_{a,h}^i \sigma_a \sigma_h\right] = 0, \quad i = \pm 1, \ldots, \pm 4$$

$$E\left[a_t y_{t+i} - \rho^i_{a,y} \sigma_a \sigma_y\right] \;=\; 0, \; i = 1, \ldots, 4.$$

Consistent with this reparameterization we redefined tests 2 through 5 so that they pertained to σ_c, σ_i, σ_h and σ_a respectively and adjusted the definitions of H1 and H4 accordingly.[17]

Figure 4 reports the small sample size of the Wald tests with asymptotic size equal to 10% (Panel A), 5% (Panel B) and 1% (Panel C) for the reparameterized system. Notice that in most cases, small sample size increases. This is true for hypotheses H1–H4, except for the test of H1 at the 1% level. For the tests based on correlations (hypotheses 6–23) this is true for 51 out of 54 cases. Notice, however, that small sample performance improves dramatically for the tests based on σ_c and σ_i over those based on σ_c/σ_y and σ_i/σ_y (hypotheses 2 & 3). Interestingly, this improvement does not translate into improved performance for the test of hypothesis H1. So while the reparameterization appears to lead to more uniform performance across the different moments, it does not solve the overall excessive small sample size of the Wald tests. And it certainly does not account for the dramatic problems associated with tests of hypothesis H4.

In the remainder of this section we discuss the factors underlying the large (small sample) size of the Wald test of hypothesis H4. For this purpose we return to the original parameterization of Ψ_2 and focus on the role played by the matrix S_T in the small sample distribution of the $\mathcal{W}$ statistic. To this end, we redid our Monte Carlo experiments using the population value of S_T, S_0, that is implied by the parameters governing the data generating process. Specifically, on each of the one thousand data sets, we estimated the parameters of the model but formed the $\mathcal{W}$ statistic using the fixed matrix S_0. We found that the $\mathcal{W}$ statistics for H4 exceed their asymptotic (1%, 5%, 10%) critical values (4%, 8%, 11%) of the time. This contrasts with our baseline findings that the $\mathcal{W}$ statistic exceeds its asymptotic (1%, 5%, 10%) critical values (37%, 51%, 58%) of the time.[18] Evidently, the fact that we must estimate S_0 accounts for a substantial part of the problem. But even when S_0 is known, relying on asymptotic distribution theory would still lead us to

---

[17]The reparameterisation indirectly affected all of the other test statistics because of the covariance between the GMM error terms.

[18]We also found that the $\mathcal{W}$ statistics for H1, H2 and H3 exceeded their asymptotic (1%, 5%, 10%) critical values (3%, 7%, 11%), (0%, 2%, 4%) and (2%, 5%, 7%) of the time, respectively. The analogous numbers in the baseline Monte Carlo where we use S_T rather than S_0 are (12%, 23%, 32%), (8%, 17%, 24%) and (7%, 13%, 20%).

22

reject hypothesis H4 too often.

A natural question arises as to whether the small sample distribution of the $\mathcal{W}$ statistic for H4 would coincide even more closely to its asymptotic distribution if we imposed the population values of D_T and $F(\Psi_T)$ as well as S_T in the Monte Carlo experiments. For our data generating process the answer is no. Indeed, we found that the small sample size of Wald test for H4 actually moved substantially farther away from its asymptotic size under these circumstances. Specifically, the $\mathcal{W}$ statistic for H4 exceeded its asymptotic (1%, 5% 10%) critical values (16%, 25%, 32%) of the time. While this does not represent a logical problem, we are surprised by the result.

Overall our results suggest that sampling error in S_T plays a substantial role in the large (small sample) sizes of Wald tests involving multiple moment conditions. This suggests an alternative way to estimate S_T. Specifically, the econometrician could calculate the implied population value of S_T for any given set of parameter estimates when estimating the model. The obvious drawback to this procedure, it that, for nontrivial models, it is computationally quite burdensome.

4 Conclusion

This paper examined the small sample properties of Generalized Method of Moments (GMM) based Wald statistics. For the data generating processes considered we found that the small sample size of these tests exceeded their asymptotic size. The problem became dramatically worse as the dimensionality of the joint tests being considered increased. We offered evidence that the basic problem has to do with the difficulty of estimating the spectral density matrix of the GMM residuals that is needed to conduct inference. Our results lead us to be very skeptical that the problem can be resolved by using any of the alternative nonparametric estimators of this matrix that have been discussed in the literature. Instead we advocate using estimators which impose as much a priori information as possible. Two important sources of such information are the economic theory being investigated and the null hypothesis being tested. There are two costs associated with pursuing this strategy. The first is computational. The second is that to pursue it the analyst will often be required to make stronger distributional assumptions about the nature of the unobservable shocks impacting on agents' environments. But, in this case, two of

the prime reasons for using a GMM strategy, as opposed to maximum likelihood methods, disappear.

References

1. Altug, Sumru (1989) "Time to Build and Aggregate Fluctuations: Some New Evidence," *International Economic Review*, Vol. 30, 889–920.

2. Andrews, Donald W.K. (1991) "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, Vol. 59, 817–858.

3. Andrews, Donald W.K. and J. Christopher Monahan (1992) "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, Vol. 60, 953–966.

4. Brayton, F. and E. Mauskopf (1985) "The MPS Model of the United States Economy," Board of Governors of the Federal Reserve Bank of Minneapolis Working Paper 425.

5. Burnside, Craig (1992) "Small Sample Properties of 2-Step Method of Moments Estimators," manuscript, University of Pittsburgh.

6. Burnside, Craig (1993) "Notes on the Linearization and GMM Estimation of Real Business Cycle Models," manuscript, University of Pittsburgh.

7. Burnside, Craig and Martin Eichenbaum (1994) "Factor Hoarding and the Propagation of Business Cycle Shocks," NBER Working Paper 4675.

8. Christiano, Lawrence J. (1988) "Why Does Inventory Investment Fluctuate So Much?," *Journal of Monetary Economics*, Vol. 21, (March/May), 247–280.

9. Christiano, Lawrence J. and Wouter den Haan (1994) "Small Sample Properties of GMM for Business Cycle Analysis," manuscript, Northwestern University.

10. Christiano, L. and Eichenbaum, M. (1992) "Current Real Business Cycle Theories and Aggregate Labor Market Fluctuations," *American Economic Review*, Vol. 82, (June), 430–450.

11. Eichenbaum, M.S. and L.P. Hansen (1990) "Estimating Models with Intertemporal Substitution Using Aggregate Time Series Data," *Journal of Business and Economic Statistics*, 8, 53–69.

12. Ferson, Wayne and Stephen E. Foerster (1991) "Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models," manuscript, University of Chicago.

13. Fuhrer, Jeffrey, George Moore and Scott Schuh (1993) "Estimating the Linear-Quadratic Inventory Model: Maximum Likelihood Versus Generalized Method of Moments," Finance and Economics Discussion Series 93-11, Board of Governors of the Federal Reserve System.

14. Gallant, A.R. (1987) *Nonlinear Statistical Models*. New York: Wiley.

15. Gregory, Allan W. and Michael R. Veall (1985) "Formulating Wald Tests of Nonlinear Restrictions," *Econometrica*, Vol. 53, 1465–70.

16. Hansen, Gary D. (1985) "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics*, Vol. 16, (November), 309–328.

17. Hansen, Lars P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, Vol. 50, 1029–1054.

18. Hansen, L.P. and K.J. Singleton (1982) "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

19. Hodrick, Robert J. and Edward C. Prescott (1980) "Post-War Business Cycle: An Empirical Investigation," manuscript, Carnegie Mellon University.

20. King, Robert G., Charles I. Plosser and Sergio Rebelo (1988) "Production, Growth and Business Cycles," *Journal of Monetary Economics*, Vol. 21, (March/May) 195–232.

21. Kocherlakota, N. (1990) "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics*, 26, 285–304.

22. Kydland, Finn E. and Edward C. Prescott (1982) "Time to Build and Aggregate Fluctuations," *Econometrica*, Vol. 50, 1435–70.

23. Leeper, Eric and Christopher S. Sims (1994) "???" manuscript, Yale University.

24. McGrattan, Ellen, Richard Rogerson and Randall Wright (1993) "Household Production and Taxation in the Stochastic Growth Model," Federal Reserve Bank of Minneapolis, Research Department, Staff Report No. 166.

25. Neely, Christopher J. (1993) "A Reconsideration of Representative Consumer Asset Pricing Models," manuscript, University of Iowa.

26. Newey, Whitney K. and Kenneth D. West (1987) "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, Vol. 55, 703-708.

27. Newey, Whitney K. and Kenneth D. West (1993) "Automatic Lag Selection in Covariance Matrix Estimation," manuscript, M.I.T.

28. Prescott, Edward C. (1986) "Theory Ahead of Business Cycle Measurement," *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol. 10, Fall.

29. Tauchen, George (1986) "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business and Economic Statistics*, Vol. 4, 397–416.

30. West, Kenneth D. and David W. Wilcox (1994) "Some Evidence on Finite Sample Distributions of Instrumental Variables Estimators of a Linear Quadratic Inventory Model," manuscript, Board of Governors of the Federal Reserve System.

TABLE 1

Small Sample Performance of Joint Tests
Using Normally Distributed White Noise Data
Estimating Standard Deviations

A. Estimated S_T, $B_T = 4$

Asymptotic Size	Small Sample Size (%)				
	$M = 1$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	2.59	3.41	6.99	16.98	58.68
5%	7.49	9.25	15.61	30.92	73.37
10%	12.65	14.93	23.32	40.10	80.29

B. Estimated S_T, $B_T = 2$

Asymptotic Size	Small Sample Size (%)				
	$M = 1$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	2.31	2.87	4.83	9.17	28.88
5%	6.90	8.26	12.22	19.91	45.62
10%	12.03	13.62	19.32	28.55	55.88

C. Estimated S_T, B_T by Andrews Procedure

Asymptotic Size	Small Sample Size (%)				
	$M = 1$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	2.27	2.91	4.71	9.06	26.64
5%	6.94	8.27	11.94	19.27	43.43
10%	11.98	13.50	19.04	27.87	53.83

D. Estimated S_T, No Lags

Asymptotic Size	Small Sample Size (%)				
	$M = 1$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	2.15	2.73	4.17	6.67	17.31
5%	6.74	7.94	10.82	16.23	32.87
10%	11.79	13.22	17.43	24.10	42.51

E. Estimated Diagonal S_T, No Lags

Asymptotic Size	Small Sample Size (%)				
	$M = 1$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	2.15	2.67	3.33	3.88	4.71
5%	6.74	7.58	9.32	11.04	13.39
10%	11.79	13.04	15.50	17.56	21.20

F. Gaussianity Applied to E

Asymptotic Size	Small Sample Size (%)				
	$M = 1$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	1.67	1.82	2.22	2.40	2.58
5%	5.94	6.08	7.20	7.72	8.53
10%	10.60	11.30	12.50	13.25	14.45

G. Impose H_0 on S_T in F

Asymptotic Size	Small Sample Size (%)				
	$M = 1$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	1.46	1.67	2.03	2.10	2.10
5%	4.61	5.33	5.97	6.58	7.26
10%	9.34	9.55	10.47	11.70	12.05

H. Impose H_0 on S_T in F, and on D_T

Asymptotic Size	Small Sample Size (%)				
	$M = 1$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	0.96	0.97	0.99	0.96	0.92
5%	5.16	4.90	5.08	5.01	4.99
10%	10.14	10.13	10.20	10.11	9.99

TABLE 2

Small Sample Performance of Joint Tests
Using Normally Distributed White Noise Data
Estimating Relative Standard Deviations

A. Estimated S_T, $B_T = 4$

Asymptotic Size	Small Sample Size (%)					
	$M = 1a$	$M = 1b$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	2.59	2.26	3.65	7.88	18.55	59.25
5%	7.49	7.09	9.55	16.62	32.30	73.11
10%	12.65	12.12	15.35	24.17	40.99	79.98

B. Estimated S_T, $B_T = 2$

Asymptotic Size	Small Sample Size (%)					
	$M = 1a$	$M = 1b$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	2.31	1.95	3.14	5.65	10.61	31.09
5%	6.90	6.63	8.58	13.08	21.73	47.21
10%	12.03	11.46	14.13	20.32	29.98	56.65

C. Estimated S_T, B_T by Andrews Procedure

Asymptotic Size	Small Sample Size (%)					
	$M = 1a$	$M = 1b$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	2.28	1.90	3.12	5.45	9.80	27.60
5%	6.88	6.57	8.46	12.87	20.65	43.72
10%	11.84	11.40	13.93	19.98	28.84	53.46

D. Estimated S_T, No Lags

Asymptotic Size	Small Sample Size (%)					
	$M = 1a$	$M = 1b$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	2.15	1.84	2.95	4.88	8.18	20.29
5%	6.74	6.42	8.09	11.90	17.92	34.91
10%	11.79	11.15	13.54	18.41	25.80	44.46

E. Impose Mutual Independence

Asymptotic Size	Small Sample Size (%)					
	$M = 1a$	$M = 1b$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	2.15	1.79	2.87	4.07	5.37	7.36
5%	6.74	6.24	7.88	10.43	12.44	16.51
10%	11.79	11.07	13.29	16.60	19.24	23.72

F. Gaussianity Applied to E

Asymptotic Size	Small Sample Size (%)					
	$M = 1a$	$M = 1b$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	1.67	1.44	2.10	2.81	3.53	4.51
5%	5.94	5.45	6.45	8.10	9.41	11.28
10%	10.60	10.25	11.68	13.48	14.81	17.33

G. Impose H_0 on S_T in F

Asymptotic Size	Small Sample Size (%)					
	$M = 1a$	$M = 1b$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	1.46	2.79	1.76	2.50	2.98	3.86
5%	4.61	6.09	5.63	7.04	8.45	9.51
10%	9.34	9.86	9.82	11.71	13.23	15.19

H. Impose H_0 on S_T in F, and on D_T

Asymptotic Size	Small Sample Size (%)					
	$M = 1a$	$M = 1b$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
1%	0.96	1.36	1.21	1.71	2.02	2.67
5%	5.16	5.43	5.36	6.19	6.71	7.76
10%	10.14	10.11	10.25	11.24	11.92	13.40

TABLE 3

Model Parameters Ψ_1
Estimates and Standard Errors[*]

Parameter	Estimate	Std. Error
θ	3.5955	(0.0377)
α	0.6422	(0.0193)
δ	0.0208	(0.0002)
γ	0.0038	(0.0012)
σ_v	0.0088	(0.0007)
g/y	0.1763	(0.0022)
g_0	1.7885	(0.0809)
g_1	-0.0019	(0.0003)
ρ	0.9456	(0.0299)
σ_ϵ	0.0152	(0.0012)
σ_ξ	0.0088	(0.0011)

[*] All standard errors shown in this table are based on estimates of S_T computed using the Bartlett window suggested by Newey and West (1987), and the automatic bandwidth selection procedure suggested by Andrews (1991).

TABLE 4

Tests of the Models*

Moment	U.S. Data	Model	W
σ_y	0.0192	0.0167	1.614
	(0.0018)	(0.0013)	(0.204)
σ_c/σ_y	0.437	0.480	2.005
	(0.029)	(0.009)	(0.157)
σ_i/σ_y	2.224	2.244	0.044
	(0.068)	(0.072)	(0.835)
σ_h/σ_y	0.859	0.795	0.990
	(0.069)	(0.051)	(0.320)
σ_h/σ_a	1.221	1.033	2.258
	(0.115)	(0.037)	(0.133)

*Numbers under the heading U.S. Data are second moments of HP-filtered U.S. data. Numbers under the heading model, are the model's implications for the corresponding moments as functions of Ψ_1. Standard errors for each are in parentheses. The p-values for the corresponding W statistics are in parentheses.

TABLE 5

Small Sample Performance of the Joint Tests

Hypothesis	Test Performed Using U.S. Data*			Size (%) of Tests[†]		
	W	p-value	MC p-value	10%	5%	1%
H1	6.64	0.25	0.48	31.7	23.0	11.9
H2	43.7	0.00	0.01	23.6	16.5	7.6
H3	35.5	0.00	0.01	20.2	13.3	6.5
H4	66.3	0.00	0.06	57.6	50.7	36.7

*The numbers under the heading 'p-value' are the p-values obtained when the W statistics for H1, H2, H3 and H4 are compared to χ^2 distributions with 5, 9, 9 and 17 degrees of freedom respectively. The numbers under the heading 'MC p-value' are obtained by comparing these statistics to the distribution of the W statistics generated by our Monte Carlo experiments.

[†]The numbers on this side of the table indicate the frequency (in %) with which the W statistics from our Monte Carlo experiments exceed the 10%, 5% and 1% critical values of the relevant χ^2 distributions.

TABLE 6

The Form of the Lag Window and Small Sample Performance*

Moment	10%			5%			1%		
	B	P	Q	B	P	Q	B	P	Q
1	16.1	16.0	15.8	10.1	9.7	9.5	3.0	3.1	3.0
2	25.5	24.0	25.4	17.9	16.3	17.5	8.2	8.0	8.3
3	25.4	25.7	25.5	18.8	19.7	19.0	8.2	9.0	8.2
4	13.2	14.2	13.3	7.3	8.0	7.3	2.0	2.2	1.8
5	12.4	13.4	12.0	6.7	7.6	6.7	1.2	1.6	1.3
6	15.5	15.0	14.9	9.0	8.9	8.8	3.2	3.1	2.6
7	14.4	14.2	14.3	8.3	8.0	7.8	2.7	2.8	2.6
8	17.1	17.4	17.6	9.7	9.5	8.6	3.0	3.0	2.6
9	18.2	18.3	17.7	10.5	10.6	9.9	2.5	3.1	2.7
10	16.3	16.9	16.4	10.5	10.5	10.1	3.8	4.4	4.0
11	7.8	8.8	7.9	3.9	5.2	4.4	0.5	1.1	0.6
12	10.6	10.1	9.3	4.0	4.5	4.0	1.3	1.4	1.2
13	13.0	12.3	12.4	6.8	6.9	6.4	2.0	2.2	2.2
14	12.8	12.8	12.5	6.9	7.5	6.7	2.8	2.8	2.4
15	17.2	16.4	16.1	10.5	9.8	9.8	4.0	3.9	3.8
16	18.9	17.6	17.9	10.7	10.1	10.2	3.6	3.7	3.3
17	19.9	19.1	19.2	11.8	11.6	10.9	3.5	3.7	3.6
18	19.1	19.3	18.6	11.1	11.1	10.4	4.0	3.8	3.5
19	12.6	13.9	12.2	6.2	6.5	5.4	1.6	1.9	1.5
20	7.1	8.5	7.4	4.8	5.5	4.3	0.5	0.9	0.7
21	11.8	11.1	10.1	6.3	5.9	5.2	1.3	1.6	1.2
22	14.2	13.6	12.4	7.2	7.1	6.9	2.2	2.4	2.3
23	14.4	13.9	13.7	8.8	8.2	8.0	2.2	2.4	2.4
H1	31.7	32.6	30.4	23.0	24.1	22.7	11.9	13.2	11.5
H2	23.6	28.1	26.1	16.5	21.3	18.8	7.6	10.6	8.3
H3	20.2	23.7	21.1	13.3	16.7	14.9	6.5	8.7	7.3
H4	57.6	63.6	59.0	50.7	56.9	52.4	36.7	43.4	38.3

*The sets of columns labelled x% refer to the sizes of tests (in %) with asymptotic size equal to x%. The labels B, P, Q refer to the Bartlett, Parzen and Quadratic Spectral windows. The Bartlett kernel columns are our baseline case, the others differ from that case only in the lag window used, and consequently in the bandwidths chosen by the Andrews (1991) procedure. Tests are numbered as described in the text.

TABLE 7

The Impact of Excluding Some Moment Restrictions*

Moment	10%		5%		1%	
	Ex	In	Ex	In	Ex	In
1	16.1	28.8	10.1	22.4	3.0	13.0
2	25.5	33.4	17.9	25.5	8.2	14.9
3	25.4	35.7	18.8	29.4	8.2	19.2
4	13.2	28.5	7.3	22.2	2.0	12.4
5	12.4	28.1	6.7	20.7	1.2	11.0
6	15.5	30.0	9.0	22.0	3.2	12.1
7	14.4	27.0	8.3	20.2	2.7	11.6
8	17.1	26.6	9.7	20.0	3.0	11.8
9	18.2	30.9	10.5	22.7	2.5	13.2
10	16.3	29.8	10.5	23.1	3.8	13.7
11	7.8	23.5	3.9	17.0	0.5	9.0
12	10.6	26.1	4.0	18.8	1.3	10.3
13	13.0	26.6	6.8	19.4	2.0	11.0
14	12.8	25.9	6.9	19.3	2.8	11.0
15	17.2	33.9	10.5	25.4	4.0	15.1
16	18.9	33.5	10.7	25.3	3.6	13.5
17	19.9	34.2	11.8	25.4	3.5	15.1
18	19.1	32.2	11.1	24.1	4.0	14.5
19	12.6	29.6	6.2	23.3	1.6	12.0
20	7.1	28.5	4.8	20.5	0.5	10.4
21	11.8	29.3	6.3	21.8	1.3	13.1
22	14.2	31.3	7.2	24.5	2.2	14.7
23	14.4	31.3	8.8	25.0	2.2	14.8
H1	31.7	85.4	23.0	82.3	11.9	71.7
H2	23.6	95.8	16.5	93.7	7.6	88.2
H3	20.2	93.5	13.3	91.6	6.5	85.9
H4	57.6	100.0	50.7	100.0	36.7	99.9

*The columns labelled 'Ex' correspond to our baseline case, while those labelled 'In' correspond to experiments in which the three moment restrictions, excluded in our baseline case, are not excluded in computing the automatic bandwidths.

33

TABLE 8

The Newey and West Automatic Bandwidth Procedure*

Moment	10%			5%			1%		
	A	NW(4)	NW(12)	A	NW(4)	NW(12)	A	NW(4)	NW(12)
1	16.1	16.5	19.0	10.1	9.8	11.3	3.0	3.3	4.0
2	25.5	27.0	24.7	17.9	18.9	17.9	8.2	9.2	9.2
3	25.4	25.1	26.8	18.8	17.6	20.5	8.2	7.5	10.0
4	13.2	12.5	16.5	7.3	7.1	9.4	2.0	2.0	2.9
5	12.4	12.1	15.8	6.7	6.7	9.4	1.2	1.3	2.8
6	15.5	16.0	18.6	9.0	9.4	10.4	3.2	3.0	4.2
7	14.4	14.5	16.1	8.3	8.7	9.3	2.7	3.4	4.2
8	17.1	16.9	18.1	9.7	9.4	11.5	3.0	2.4	3.2
9	18.2	17.5	19.0	10.5	9.6	11.3	2.5	2.7	3.8
10	16.3	17.2	18.4	10.5	10.0	12.5	3.8	3.9	5.6
11	7.8	7.7	11.0	3.9	3.8	6.0	0.5	0.5	1.8
12	10.6	11.1	14.0	4.0	5.0	7.9	1.3	1.5	2.2
13	13.0	13.1	14.7	6.8	7.2	9.1	2.0	2.3	3.3
14	12.8	13.8	15.3	6.9	7.6	9.2	2.8	2.9	4.2
15	17.2	17.7	21.0	10.5	10.5	13.0	4.0	4.0	4.7
16	18.9	18.5	21.5	10.7	11.3	13.1	3.6	4.1	5.3
17	19.9	19.6	22.7	11.8	11.8	14.0	3.5	4.2	5.3
18	19.1	19.8	21.2	11.1	11.2	13.0	4.0	3.8	5.0
19	12.6	11.3	16.1	6.2	6.1	8.8	1.6	1.3	2.6
20	7.1	7.5	11.7	4.8	4.5	6.7	0.5	0.6	1.7
21	11.8	11.8	15.8	6.3	6.6	9.1	1.3	1.4	2.8
22	14.2	13.7	17.9	7.2	8.1	11.1	2.2	2.6	4.6
23	14.4	14.9	18.4	8.8	9.4	12.1	2.2	2.9	4.4
H1	31.7	31.8	42.4	23.0	21.8	34.6	11.9	12.1	20.8
H2	23.6	24.8	41.8	16.5	17.2	34.8	7.6	8.7	23.9
H3	20.2	20.7	38.6	13.3	14.9	30.9	6.5	8.4	20.7
H4	57.6	55.9	78.8	50.7	49.8	73.0	36.7	36.4	63.0

*The columns labelled 'A' correspond to our baseline case, which uses the Andrews automatic bandwidth procedure, while those labelled 'NW(x)' correspond to experiments using the Newey and West procedure, with x being the value of n.

TABLE 9

Variable Versus Fixed Bandwidth[*]

| Moment	10%			5%			1%		
	V	2	4	V	2	4	V	2	4
1	16.1	16.4	16.6	10.1	10.5	9.8	3.0	3.1	3.0
2	25.5	28.3	23.9	17.9	20.1	16.3	8.2	9.6	7.9
3	25.4	24.2	25.6	18.8	17.4	19.6	8.2	6.4	8.7
4	13.2	12.0	13.6	7.3	6.5	8.0	2.0	1.1	2.1
5	12.4	10.9	13.6	6.7	5.9	7.5	1.2	1.1	1.6
6	15.5	15.5	15.3	9.0	9.3	8.8	3.2	2.6	3.2
7	14.4	14.8	14.4	8.3	8.5	8.0	2.7	2.4	2.9
8	17.1	17.6	17.1	9.7	8.9	10.3	3.0	2.2	3.2
9	18.2	17.8	17.9	10.5	9.5	10.4	2.5	2.2	2.9
10	16.3	15.8	17.4	10.5	10.1	10.4	3.8	3.5	4.3
11	7.8	6.1	8.7	3.9	2.9	4.4	0.5	0.3	0.9
12	10.6	9.2	11.4	4.0	3.7	4.7	1.3	1.2	1.4
13	13.0	12.8	12.7	6.8	6.8	7.0	2.0	2.4	2.2
14	12.8	13.1	13.0	6.9	7.3	7.5	2.8	2.5	3.1
15	17.2	17.1	16.5	10.5	11.0	9.8	4.0	3.9	3.6
16	18.9	18.7	18.4	10.7	10.9	10.7	3.6	3.9	3.9
17	19.9	19.8	20.1	11.8	11.8	11.8	3.5	3.4	4.2
18	19.1	18.8	19.5	11.1	10.4	11.7	4.0	3.2	3.8
19	12.6	9.5	13.6	6.2	4.1	6.1	1.6	1.0	1.7
20	7.1	5.8	8.7	4.8	2.5	4.9	0.5	0.2	0.5
21	11.8	9.7	12.7	6.3	5.0	6.9	1.3	0.9	1.5
22	14.2	13.1	14.4	7.2	7.2	7.6	2.2	2.4	2.7
23	14.4	14.2	14.2	8.8	8.4	8.6	2.2	2.4	2.3
H1	31.7	28.6	32.3	23.0	19.6	24.3	11.9	10.7	12.8
H2	23.6	20.0	25.6	16.5	12.9	19.4	7.6	5.8	8.7
H3	20.2	16.7	21.6	13.3	10.6	14.9	6.5	5.3	7.2
H4	57.6	49.6	62.7	50.7	41.2	55.2	36.7	27.5	41.6

[*]The sets of columns labelled $x\%$ refer to tests with asymptotic size equal to $x\%$. The labels V, 2 and 4 refer to variable bandwidths picked with the Andrews (1991) procedure, a fixed bandwidth of 2 and a fixed bandwidth of 4 respectively. All results are based on our other baseline choices.

TABLE 10

Variable Versus Fixed Bandwidth[*]

Moment	10%			5%			1%		
	V	6	8	V	6	8	V	6	8
1	16.1	17.8	19.0	10.1	10.4	11.6	3.0	3.6	3.9
2	25.5	23.1	23.2	17.9	15.8	16.6	8.2	7.6	8.0
3	25.4	26.7	26.5	18.8	20.0	20.8	8.2	9.8	9.9
4	13.2	16.0	16.9	7.3	9.0	10.3	2.0	2.5	2.8
5	12.4	15.1	16.6	6.7	8.6	9.6	1.2	2.2	2.8
6	15.5	16.3	17.8	9.0	9.3	9.9	3.2	4.2	5.0
7	14.4	15.5	15.9	8.3	8.6	9.9	2.7	3.4	3.6
8	17.1	17.3	17.6	9.7	10.7	11.2	3.0	3.5	4.0
9	18.2	18.0	19.0	10.5	11.0	11.6	2.5	3.4	4.0
10	16.3	18.6	19.5	10.5	11.5	12.5	3.8	5.0	5.3
11	7.8	11.0	12.2	3.9	5.7	7.0	0.5	1.4	1.7
12	10.6	13.3	14.6	4.0	5.9	7.4	1.3	2.1	2.3
13	13.0	13.6	14.4	6.8	8.1	9.1	2.0	3.0	3.3
14	12.8	14.2	15.2	6.9	8.7	9.4	2.8	3.7	4.1
15	17.2	18.2	20.2	10.5	10.8	12.6	4.0	4.0	4.9
16	18.9	20.9	23.0	10.7	12.0	14.0	3.6	4.4	5.3
17	19.9	21.0	22.7	11.8	13.1	14.8	3.5	4.8	5.0
18	19.1	21.0	21.8	11.1	12.4	13.9	4.0	5.1	5.2
19	12.6	15.1	17.2	6.2	8.3	10.1	1.6	2.4	3.3
20	7.1	10.9	12.6	4.8	6.0	7.3	0.5	1.6	2.1
21	11.8	15.2	17.1	6.3	8.1	9.3	1.3	2.2	2.7
22	14.2	17.3	18.8	7.2	10.0	12.4	2.2	3.5	4.8
23	14.4	16.8	19.2	8.8	10.9	12.5	2.2	3.6	5.1
H1	31.7	37.8	42.9	23.0	28.5	34.1	11.9	16.3	20.3
H2	23.6	34.7	43.2	16.5	26.4	35.6	7.6	14.8	21.6
H3	20.2	29.8	39.7	13.3	22.0	30.8	6.5	11.8	17.8
H4	57.6	76.3	86.6	50.7	69.6	81.9	36.7	55.9	72.0

[*]The sets of columns labelled x% refer to tests with asymptotic size equal to x%. The labels V, 6 and 8 refer to variable bandwidths picked with the Andrews (1991) procedure, a fixed bandwidth of 6 and a fixed bandwidth of 8 respectively. All results are based on our other baseline choices.
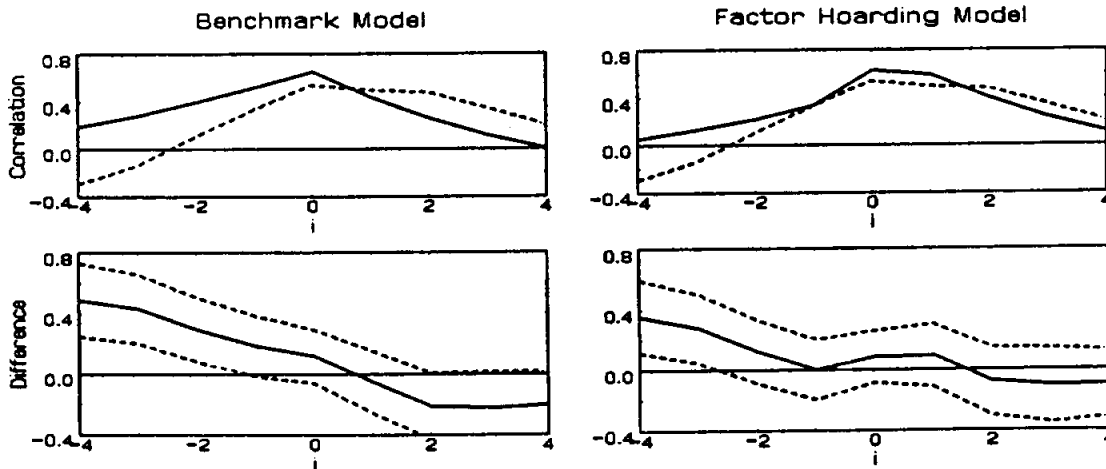
TABLE 11

First Order VAR Prewhitening*

Moment	10%		5%		1%	
	Ex	In	Ex	In	Ex	In
1	7.1	11.5	4.2	7.7	1.2	3.4
2	8.5	14.1	5.8	9.5	2.0	4.6
3	15.2	21.6	9.2	15.4	3.2	6.9
4	6.6	12.7	4.0	9.2	1.0	3.3
5	7.3	11.9	4.1	7.8	1.3	3.7
6	4.6	11.0	2.5	6.3	1.1	2.9
7	5.1	9.7	3.0	6.0	1.2	3.3
8	5.5	10.6	3.2	6.5	0.5	2.7
9	7.0	12.3	3.7	7.5	1.0	2.7
10	7.0	11.6	3.9	7.7	1.4	4.2
11	2.8	6.7	1.8	3.9	0.3	1.8
12	3.6	6.9	1.8	3.9	0.6	1.8
13	3.0	6.5	1.7	4.1	0.4	1.1
14	3.6	6.8	1.9	4.4	0.7	2.5
15	6.6	12.4	4.2	8.5	1.5	3.2
16	8.5	15.2	4.9	10.4	1.7	4.2
17	8.4	13.8	4.5	9.5	1.4	4.6
18	8.6	14.5	4.3	9.9	1.6	4.3
19	7.3	13.3	3.2	9.0	0.9	3.3
20	4.1	7.7	2.3	4.5	0.8	2.4
21	3.7	6.9	2.2	5.2	0.6	2.0
22	2.6	6.4	1.4	3.6	0.5	1.8
23	4.6	7.5	2.9	4.5	0.7	2.4
H1	39.8	68.5	30.8	62.9	19.7	51.2
H2	49.5	89.7	42.3	85.9	28.9	79.9
H3	49.1	88.8	40.5	84.5	28.1	77.2
H4	91.1	100.0	88.1	100.0	81.8	99.7

*The columns labelled 'Ex' correspond to our baseline case, while those labelled 'In' correspond to experiments in which the three moment restrictions, excluded in our baseline case, are not excluded in computing the automatic bandwidths.

37

FIGURE 1

CORRELATION OF APL_t WITH H_{t+i} (HP FILTER)*

Benchmark Model

Factor Hoarding Model
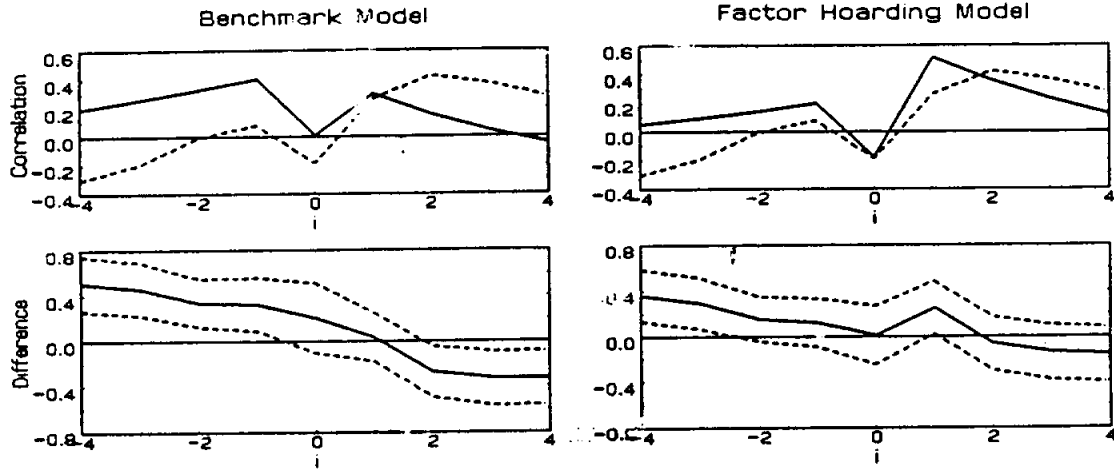
FIGURE 2

CORRELATION OF APL_t WITH Y_{t+i} (HP FILTER)*

Benchmark Model

Factor Hoarding Model

*In the Correlation panels: solid line – model predicted correlations, dashed line – sample correlations. In the Difference panels the dashed lines represent a 2-standard error band around the difference.
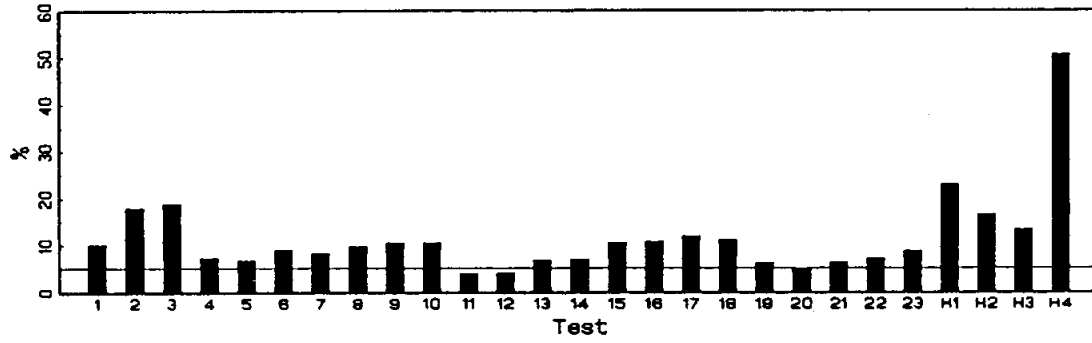
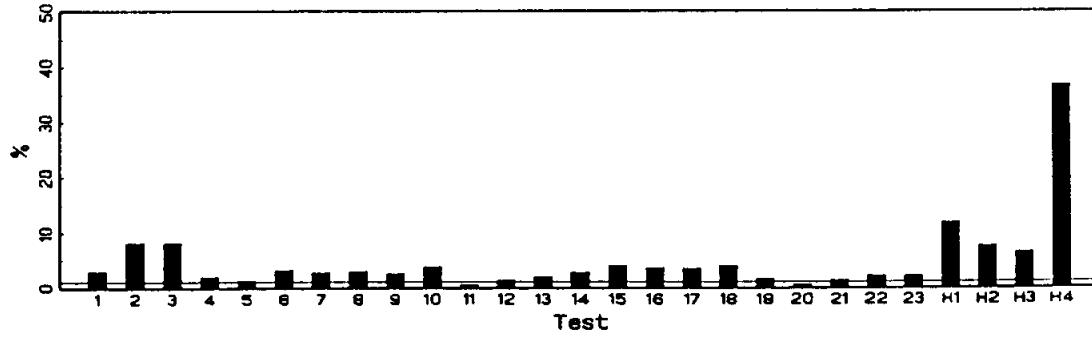# FIGURE 3

SMALL SAMPLE SIZE OF THE W TESTS
RBC Example

A. Asymptotic Size = 10%
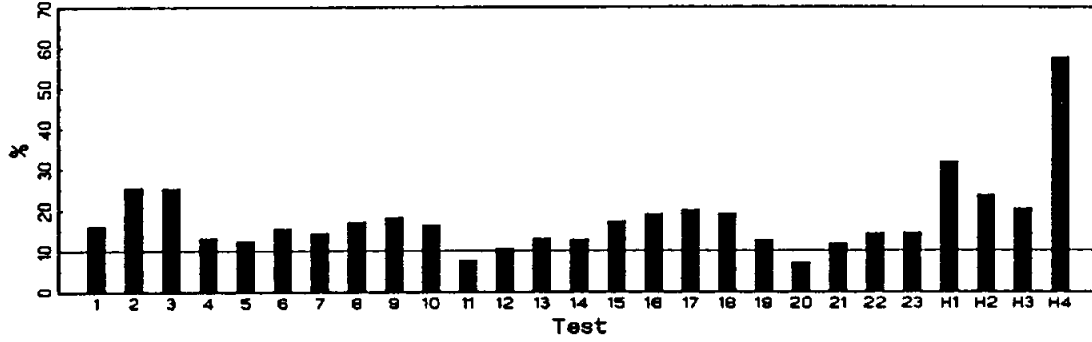
B. Asymptotic Size = 5%

C. Asymptotic Size = 1%

FIGURE 4

Small Sample Size of the $\mathcal{W}$ Tests
Reparameterized RBC Example

A. Asymptotic Size = 10%
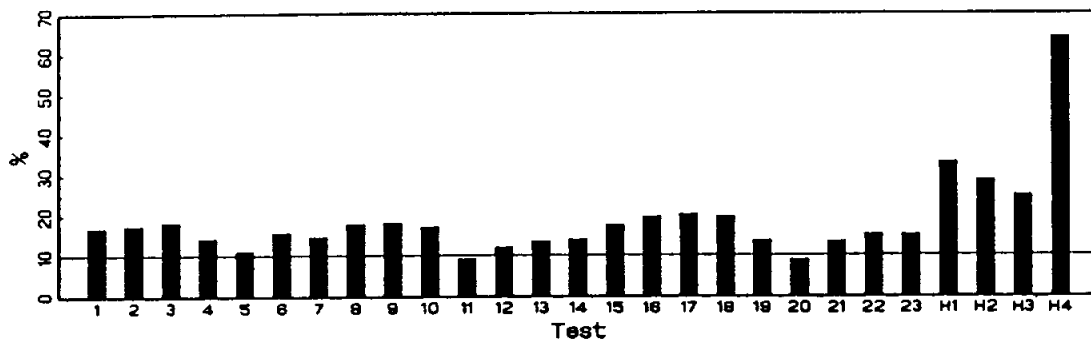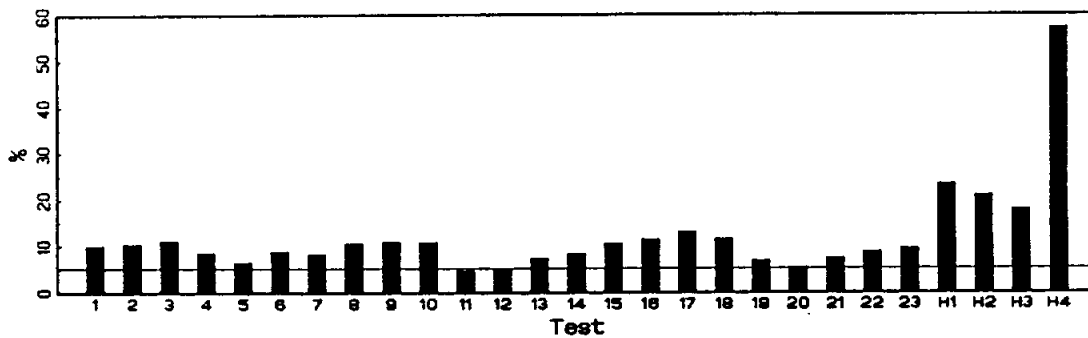
B. Asymptotic Size = 5%

C. Asymptotic Size = 1%
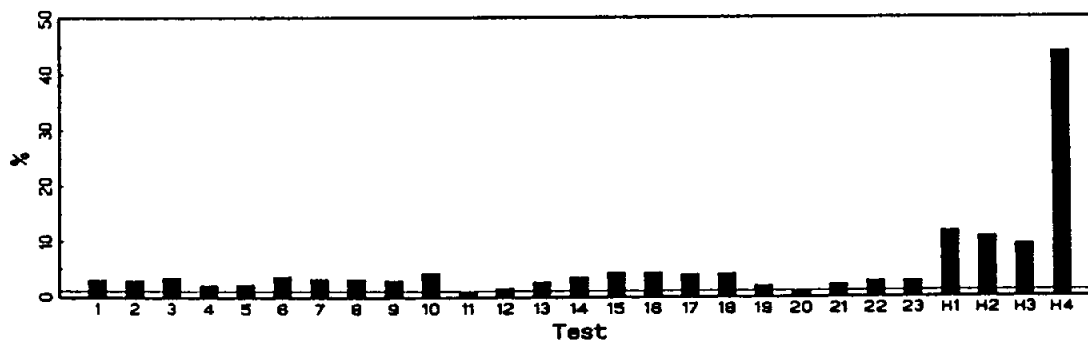
A Covariance Matrices for the Reparametrized White Noise Case

In this appendix we discuss our estimators of the asymptotic variance-covariance matrix of $\hat\theta$. The estimators S_T^1, S_T^2 and S_T^3 are defined as for $\hat\sigma$. When we allow the econometrician to exploit the lack of serial correlation in the data generating mechanism, we obtain

- S_T^4, with $(1,1)$ element $\frac{1}{T}[\sum_{t=1}^T (X_{1t}^2 - \hat\sigma_1^2)^2]$, $(1,j)$ and $(j,1)$ elements $\frac{1}{T}[\sum_{t=1}^T (X_{1t}^2 - \hat\sigma_1^2)(X_{jt}^2 - \hat\theta_j^2 X_{1t}^2)]$ for $j \geq 2$, and (i,j) element given by $\frac{1}{T}[\sum_{t=1}^T (X_{it}^2 - \hat\theta_i^2 X_{1t}^2)(X_{jt}^2 - \hat\theta_j^2 X_{1t}^2)]$, for $i \geq 2$, and $j \geq 2$.

If the econometrician imposes the mutual independence restriction then we obtain

- S_T^5, with $(1,1)$ element $\frac{1}{T}\sum_{t=1}^T X_{1t}^4 - \hat\sigma_1^4$, $(1,j)$ and $(j,1)$ elements $\hat\theta_j^2\left(\hat\sigma_1^4 - \frac{1}{T}\sum_{t=1}^T X_{1t}^4\right)$ for $j \geq 2$, (i,i) element $\frac{1}{T}\sum_{t=1}^T X_{it}^4 - \hat\theta_i^4\hat\sigma_1^4 + \hat\theta_i^4\left(\frac{1}{T}\sum_{t=1}^T X_{1t}^4 - \hat\sigma_1^4\right)$ for $i \geq 2$ and (i,j) element given by $\hat\theta_i^2\hat\theta_j^2\left(\frac{1}{T}\sum_{t=1}^T X_{1t}^4 - \hat\sigma_1^4\right)$ for $i \geq 2$ and $2 \leq j \neq i$.

If in addition the econometrician exploits the fact that the X_{it} are Gaussian, we have that $E(X_{1t}^4) = 3\sigma_1^4$ and $E(X_{jt}^4) = 3\sigma_j^4 = 3\theta_j^4\sigma_1^4$ for $j \geq 2$. This restriction can be imposed on S_T^5, which yields

- S_T^6, with $(1,1)$ element $2\hat\sigma_1^4$, $(1,j)$ and $(j,1)$ elements $-2\hat\theta_j^2\hat\sigma_1^4$ for $j \geq 2$, (i,i) element $4\hat\theta_i^4\hat\sigma_1^4$ for $i \geq 2$ and (i,j) element given by $2\hat\theta_i^2\hat\theta_j^2\hat\sigma_1^4$ for $i \geq 2$ and $2 \leq j \neq i$.

We consider two types of hypotheses when estimating $\hat\theta$: (i) hypothesis $\tilde H_M : \theta_i = 1$, $i = 1,\ldots,M$, and (ii) hypothesis $\tilde H_{1b} : \theta_2 = 1$. When we impose $\tilde H_M$ we obtain

- S_T^7, with $(1,1)$ element given by 2; $(1,j)$ and $(j,1)$ element -2 for $2 \leq j \leq M$ and $-2\hat\theta_j^2$ for $j > M$; (i,i) element equal to 4 for $2 \leq i \leq M$ and $4\hat\theta_i^4$ for $i > M$; and (i,j) element equal to 2 for $2 \leq i,j \leq M$, $2\hat\theta_i^2$ for $2 \leq i \leq M$, $j > M$ and $2\hat\theta_i^2\hat\theta_j^2$ for $i,j > M$.

When we impose hypothesis $\tilde H_{1b}$ we obtain

- S_T^7, which is identical to S_T^6 except in the second row and second column. The diagonal $(2,2)$ element is $4\hat\sigma_1^4$ while the $(2,j)$ and $(j,2)$ elements are $2\hat\theta_j^2\hat\sigma_1^4$, $j > 2$.

For each of the estimators above we construct an estimator for the variance-covariance matrix of the GMM estimator:

$$\hat V_T^k = [D_T^{1\prime}(S_T^k)^{-1}D_T^1]^{-1},$$

where D_T^1 is a diagonal matrix with $(1,1)$ element $-2\hat\sigma_1$ and (i,i) element $-2\hat\theta_i\hat\sigma_1^2$, for $j \geq 2$. Since the null hypothesis $\tilde H_M$ can also be imposed on D_T^1, we also consider the estimator

$$\hat V_T^8 = [D_T^{2\prime}(S_T^7)^{-1}D_T^2]^{-1},$$

where D_T^2 is a diagonal matrix with iith element -2 for $i \leq M$, and $-2\hat\theta_i$ for $i \geq M$. For the hypothesis $\tilde H_{1b}$ we have D_T^2 equal to D_T^1 except that the $(2,2)$ element is $-2\hat\sigma_1^2$.

B   The Euler Equation for Capital in the RBC Example

In this appendix we discuss our procedure for ensuring that the Euler equation for capital holds exactly in the data generating process underlying our Monte Carlo experiments. The Euler equation for K_{t+1} does not hold exactly for our linearized representation of the model. This equation is given by

$$E\left[1 - \beta\left(\frac{C_t}{C_{t+1}}\right)\left((1-\alpha)(1-\phi^{-1})\frac{Y_{t+1}}{K_{t+1}} + 1\right)\right] = 0. \tag{B.1}$$

As a result, when we estimate the model using artificially generated time series from the linearized model it is important to adjust this moment restriction appropriately. We compute the expectation in equation (B.1) for our linearized model (it is approximately 2×10^{-5}) evaluated at the parameter values we use to generate the artificial data. We then center the moment restriction around that value rather than 0. This expectation, denoted by ε, is computed as follows

$$
\begin{aligned}
\varepsilon &= E\left[1 - \beta\left(\frac{C_t}{C_{t+1}}\right)\left((1-\alpha)(1-\phi^{-1})\frac{Y_{t+1}}{K_{t+1}} + 1\right)\right] \\
&= E\left[1 - \beta\exp(c_t - c_{t+1})\left((1-\alpha)(1-\phi^{-1})\exp(y_{t+1} - k_{t+1}) + \exp(-\gamma - v_{t+1})\right)\right],
\end{aligned}
$$

Let $s_t = \begin{pmatrix} 1 & k_t & N_t & v_t & g_t & \xi_t \end{pmatrix}'$ and $\tilde{s}_{t+1} = \begin{pmatrix} s'_{t+1} & s'_t \end{pmatrix}'$. Any variable in the linearized model, say z_t, determined at time t is given as a function $\pi'_z s_t$, for some vector π_z determined by the solution to the model. Therefore we can write the Euler equation error simply as

$$\varepsilon = E\left[1 - \beta\left((1-\alpha)(1-\phi^{-1})\exp(\mu'_1 \tilde{s}_{t+1}) + \exp(-\gamma)\exp(\mu'_2 \tilde{s}_{t+1})\right)\right],$$

where

$$\mu_1 = \begin{pmatrix} -\pi_c + \pi_y \\ \pi_c - \pi_k \end{pmatrix} \qquad \mu_2 = \begin{pmatrix} -\pi_c - \pi_v \\ \pi_c \end{pmatrix}.$$

In our simulations we assume that the innovations to the exogenous variables are normally distributed. In this case the properties of log-normal random variables can be exploited to show that

$$\varepsilon = 1 - \beta\left[(1-\alpha)(1-\phi^{-1})\exp(\mu'_1 E\tilde{s} + \frac{1}{2}\mu'_1 \tilde{\Gamma}_s \mu_1) + \exp(-\gamma + \mu'_2 E\tilde{s} + \frac{1}{2}\mu'_2 \tilde{\Gamma}_s \mu_2)\right],$$

where $E\tilde{s}$ and $\tilde{\Gamma}_s$ are the mean and unconditional covariance matrix of $\tilde{s}_{t+1}$. These are both computable as a by-product of the solution method.

C Alternative Covariance Matrix Estimators

This appendix considers the robustness of the results presented in section 3 to alternative estimators, S_T, of the the weighting matrix S_0.

We consider various forms of S_T which depend on

   i. whether we include a small sample correction or not,

   ii. the form of the lag window, $k(\cdot)$,

   iii. the method for determining the bandwidth, B_T, and

   iv. whether we prewhiten the errors $\hat{u}_t$.

As described in the text, we take as our baseline case an estimator which

   i. does not apply any small sample correction,

   ii. uses the Bartlett lag window suggested by Newey and West (1987), and

   iii. selects the bandwidth automatically using the method suggested by Andrews (1991), and

   iv. does not prewhiten the errors.

C.1 The Small Sample Correction

The large sample performance of the tests is unaffected by the inclusion of a small sample degrees of freedom correction in the estimator S_T. In a regression context, Andrews (1991) suggests the small sample correction of multiplying S_T by a factor of $T/(T-d)$ where d is the dimension of the coefficient vector. In our simulated samples, the sample size is $T = 113$, while the length of the parameter vector Ψ is $d = 27$. Therefore, applying Andrews' small sample correction increases the magnitude of the elements of S_T by a factor of 31%. Since the effect is uniform, applying the correction unambiguously decreases all test statistics by 31%. This decreases the small sample size of the tests, although this effect will not be uniform. We did not apply the small sample correction in our baseline experiments. Although applying the correction would have improved our results somewhat, we found that this was special to results based on the HP filter, as opposed to results based on first differenced data.

C.2 The Lag Window

We consider three forms of lag window, $k(\cdot)$. Newey and West (1987) suggest using the Bartlett kernel given by

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

Gallant (1987) proposes using the Parzen kernel given by

$$k_P(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \le |x| \le 1/2, \\ 2(1 - |x|)^3 & \text{for } 1/2 \le |x| \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Andrews (1991) examines the properties of the Quadratic Spectral (QS) kernel given by

$$k_Q(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right).$$

Andrews (1991) shows that within a certain class of estimators, which includes each of these kernels, the QS kernel is optimal in the sense that it minimizes the asymptotic MSE of S_T.

We use the Bartlett kernel as our baseline lag window. Holding the other elements of the baseline estimator fixed we examine the small sample performance of our Wald tests using the Parzen and QS kernels. The results are summarized in Table 6 which shows the small sample size of the tests using the different lag windows. The results indicate that the small sample size of the tests is insensitive to the choice of lag window, at least in our example.

C.3 The Choice of Bandwidth

An issue which arises immediately with these estimators is how to choose the bandwidth parameter B_T for a given kernel. Andrews (1991) shows that the optimal bandwidths (in an MSE sense) for the three kernels are

$$B_T^* = \begin{cases} 1.1447[\alpha(1)T]^{1/3} & \text{Bartlett kernel} \\ 2.6614[\alpha(2)T]^{1/5} & \text{Parzen kernel} \\ 1.3221[\alpha(2)T]^{1/5} & \text{QS kernel,} \end{cases}$$

where $\alpha(q)$ is a function which depends on the unknown spectral density matrix of u_t given by

$$f(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\Omega_j e^{-ij\lambda},$$

where $\Omega_j = Eu_t(\Psi_0)u_{t-j}(\Psi_0)'$.

Since S_T is a matrix estimator, its MSE is typically measured with respect to some weighting scheme such as (following Andrews 1991)

$$\text{MSE}(T/B_T, S_T, W) = \frac{T}{B_T}E\text{vec}(S_T - S_0)'W\text{vec}(S_T - S_0),$$

where W is some $d^2 \times d^2$ positive definite matrix. The measure of MSE depends on the choice of the matrix W. Given a particular matrix W the optimal bandwidth formulas can be made operational since

$$\alpha(q) = \frac{2(\text{vec}\, f^{(q)})'W\text{vec}\, f^{(q)}}{\text{tr}W(I + K_{pp})f(0) \otimes f(0)},$$

44

where K_{pp} is defined so that $\text{vec}(A') = K_{pp}\text{vec}(A)$, and

$$f^{(q)} = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} |j|^q \Omega_j.$$

Automated bandwidth selection procedures provide a means of estimating the α's in the above formulas.

C.3.1 Andrews' (1991) Automated Bandwidth Procedures

Andrews (1991) proposes various automatic bandwidth estimators. These are data-based procedures which implement the above formulas for estimates of $\alpha(1)$ and $\alpha(2)$. There are many possible procedures, both parametric and nonparametric, that can be used to estimate $\alpha(1)$ and $\alpha(2)$. Parametric estimators require choosing an approximating parametric model for the errors u_t. Typical choices are parsimoniously parameterized and may model the errors individually. They further require the choice of a weighting matrix W.

Since the possibilities were numerous, we chose perhaps the simplest approach which is to choose a weighting matrix which only puts weight on the diagonal elements of S_T and to model each error term as an AR(1). Of course, the errors do not follow AR(1) processes but this does not affect consistency of the estimators for S_0. Rather it generates a bias in estimates of the optimal bandwidth. In the AR(1) case

$$\hat{\alpha}(2) = \sum_{i=1}^{k} w_i \frac{4\hat{\rho}_i^2\hat{\sigma}_i^4}{(1-\hat{\rho}_i)^8} \Big/ \sum_{i=1}^{k} w_i \frac{\hat{\sigma}_i^4}{(1-\hat{\rho}_i)^4}$$

$$\hat{\alpha}(1) = \sum_{i=1}^{k} w_i \frac{4\hat{\rho}_i^2\hat{\sigma}_i^4}{(1-\hat{\rho}_i)^6(1+\hat{\rho}_i)^2} \Big/ \sum_{i=1}^{k} w_i \frac{\hat{\sigma}_i^4}{(1-\hat{\rho}_i)^4},$$

where w_i is the weight given to error term i in computing the estimator, and $(\hat{\rho}_i, \hat{\sigma}_i)$ are standard estimates of the parameters of the AR model obtained from residual i. The simplest estimator sets $w_i = 1$ for all i.

Andrews suggests setting w_i to zero for any error terms corresponding to a constant regressor in a regression model. Presumably, this is motivated by the fact that the covariance properties of those error terms are qualitatively dissimilar to the covariance properties of the error terms corresponding to nonconstant regressors. In our examples, we placed no weight on the error terms corresponding to (17), (18) and (22) and unit weight on all other error terms, as these error terms behave very differently than the others. This constitutes our baseline method. To assess the impact of excluding these three moment restrictions we compared our baseline results to experiments where they were not excluded. In our baseline experiments the median bandwidth from the 1000 draws was 2.78 for the Bartlett kernel. With equal weight given to all moment conditions the median bandwidth rose significantly to 40.1. Furthermore, as can be seen in Table 7, the small sample performance of some of our tests deteriorated significantly.

45

C.3.2 Newey and West's (1993) Automated Bandwidth Procedure

Newey and West's (1993) procedure is related to the procedures outlined above but is nonparametric in the sense that no pseudo-model of the residuals is specified in order to estimate the α's. Newey and West note that when the MSE criterion is rewritten as $\tilde{w}'(S_T - S_0)\tilde{w}$ for some $d \times 1$ vector $\tilde{w}$, the formula for $\alpha(q)$ can be rewritten as

$$\alpha(q) = \left[\sum_{j=-\infty}^{\infty} |j|^q \tilde{w}' \Omega_j \tilde{w} \right]^2 \bigg/ \left[\sum_{j=-\infty}^{\infty} \tilde{w}' \Omega_j \tilde{w} \right]^2.$$

In order to estimate $\alpha(q)$, Newey and West (1993) suggest the approximation

$$\hat{\alpha}(q) = \left[\sum_{j=-n}^{n} |j|^q \tilde{w}' \hat{\Omega}_j \tilde{w} \right]^2 \bigg/ \left[\sum_{j=-n}^{n} \tilde{w}' \hat{\Omega}_j \tilde{w} \right]^2,$$

where n is chosen *a priori* in order to be consistent with $n \to \infty$ and $n/T^{2/9} \to \infty$ (for the Bartlett window) as $T \to \infty$. Newey and West cite evidence that S_T is less sensitive to arbitrary n than it is to arbitrary choices of B_T.

We present results for choices of $n = 4$ and $n = 12$. The weight vector we use puts zero weight on the same moment restrictions we excluded from our baseline Andrews method. The results are summarized in Table 8. When we chose $n = 4$ we obtained very similar results to when we used the Andrews procedure. This is not surprising: the median bandwidth chosen by the Newey and West procedure was 2.74 while the median bandwidth chosen by the Andrews procedure was 2.78. When we used $n = 12$ this increased the median bandwidth of the Newey and West procedure to 7.39. This led to massive overrejections of the joint hypotheses similar in scale to when we used a fixed bandwidth of 8 (see the next subsection). Overall our results indicate that automated procedures may perform similarly but only if they are 'tuned' in a way that happens to lead to similar bandwidths. We suspect that the Andrews procedure, while it has no parameter like n to be chosen, would be analagously sensitive to the choice of pseudo-model for the error terms.

C.3.3 Arbitrary Fixed Bandwidth

It is difficult to compare results obtained with fixed bandwidths to those obtained using variable bandwidths, since any results we find may not be interpretable beyond the confines of our example. However, in Table 9 and Table 10 we compare our baseline results with results obtained using fixed bandwidths of 2, 4, 6 and 8 in repeated samples. While the results are mixed for small bandwidths, the results indicate a deterioration of small sample performance (especially for joint tests) for bandwidths of 6 and 8. What is also clear from these tables is that the bandwidth affects the various tests differentially.

C.4 Prewhitening of the Errors

Andrews and Monahan (1992) suggest a procedure which prewhitens the error terms as a first step prior to the computation of S_T. Prewhitening is motivated by the apparent

problem in estimating S_T when the nature of the persistence in the errors is unknown. A particular bandwidth in tandem with a particular lag window may not adequately capture the nature of the persistence in the errors in small samples. The whiter are the error terms the less important will be the choice of lag window and bandwidth. A prewhitening procedure uses an arbitrary procedure to whiten the error terms, computes the equivalent of S_T for those whiter errors, then recolors the estimated matrix.

As an example, suppose that a first-order VAR is fit to process $\hat{u}_t$,

$$\hat{u}_t = \hat{\Pi}\hat{u}_{t-1} + \hat{\eta}_t.$$

Suppose that $\hat{\Pi}$ converges to Π asymptotically so that the errors $u_t(\Psi_0)$ have the representation

$$u_t(\Psi_0) = \Pi u_{t-1}(\Psi_0) + \eta_t.$$

Define

$$S_0^\eta = \sum_{j=-\infty}^{\infty} E(\eta_t \eta_{t-j}') = \sum_{j=-\infty}^{\infty} \Omega_j^\eta.$$

Then notice that

$$S_0 = (I - \Pi)^{-1} S_0^\eta (I - \Pi')^{-1}.$$

An analagous estimator S_T is

$$S_T = (I - \hat{\Pi})^{-1} S_T^\eta (I - \hat{\Pi}')^{-1},$$

where S_T^η is an estimator of the variety described in previous sections applied to $\hat{\eta}_t$. For higher order VARs represented as $[I - \Pi(L)]u_t = \eta_t$, the corresponding estimator would be

$$S_T = \left[I - \hat{\Pi}(I)\right]^{-1} S_T^\eta \left[I - \hat{\Pi}(I)'\right]^{-1}.$$

We conducted experiments using 1st-order VARs for the error terms. The results of these experiments are presented in Table 11. Given that we prewhitened the errors we thought that comparisons should be made with both the 'Ex' and the 'In' columns in Table 7. Notice that small sample sample performance of the tests changes dramatically. Some of the tests reject less often, but the joint tests perform terribly. The test of H4 almost always rejects in the 'In' case where we include all the error terms in our bandwidth calculations. These results are somewhat surprising. We might expect prewhitening to improve performance. However, the median bandwidth chosen by the automated procedures rises to 7.72 in the 'Ex' case and drops to 29.7 in the 'In' case. The rise in the 'Ex' case is not surprising since the included errors have been projected onto lags of the excluded errors.

D   Data

In this appendix we describe the data that was used to estimate the RBC model of section 3. Private consumption, C_t, was measured as the sum of private sector expenditures on nondurable goods plus services plus the imputed service flow from the stock of durable

goods. The first two measures were obtained from the Survey of Current Business. The third measure was obtained from Brayton and Mauskopf (1985). Government consumption, G_t, was measured by real government (federal, state and local) purchases of goods minus real government investment. The government data was provided to us by John Musgrave at the Bureau of Economic Analysis. The official capital stock, $\tilde{K}_t$, was measured as the sum of consumer durables, producer structures and equipment, and government and private residential capital plus government non-residential capital. Data on gross investment, I_t, are the flow data that conceptually match the capital stock data. Gross output, Y_t, was measured as $(C_t + G_t + I_t)$ plus time t inventory investment. Our basic measure of hours worked is the quarterly time series constructed by Hansen (1985), which we refer to as household hours. The data cover the period 1955:3-1984:1 and were converted to per capita terms using an efficiency weighted measure of the population.[19] We use Prescott's (1986) model of measurement error in hours worked. In particular we assume that the log of measured hours worked differs from the log of actual hours worked by an i.i.d. random variable that has mean zero and standard deviation σ_ξ. To estimate σ_ξ we need two measures of hours worked. The first is Hansen's measure of hours worked which is based on the household survey conducted by the Bureau of the Census. The second is the establishment survey conducted by the Bureau of Labor Statistics.

---

[19]See Christiano (1988, appendix) for further details.