Federal Reserve Bank of Minneapolis
Research Department Staff Report 199

September 1995

Small Sample Properties of GMM for Business Cycle Analysis

Lawrence J. Christiano*

Federal Reserve Bank of Minneapolis,
Northwestern University,
National Bureau of Economic Research,
and Federal Reserve Bank of Chicago

Wouter den Haan*

University of California, San Diego

ABSTRACT

We investigate, by Monte Carlo methods, the finite sample properties of GMM procedures for conducting inference about statistics that are of interest in the business cycle literature. These statistics include the second moments of data filtered using the first difference and Hodrick-Prescott filters, and they include statistics for evaluating model fit. Our results indicate that, for the procedures considered, the existing asymptotic theory is not a good guide in a sample the size of quarterly postwar U.S. data.

*The authors are grateful to Neil Ericsson, Rob Engle, Andrew Levin, Masao Ogaki, Kenneth West, and the referees for comments, and Christiano is grateful to the National Science Foundation for financial support. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Introduction

Statistical tools based on the generalized method of moments (GMM) procedures outlined by Hansen (1982) are increasingly being used in the analysis of business cycles. (See, for example, Backus, Gregory, and Zin 1989; Backus and Kehoe 1992; Backus, Kehoe, and Kydland 1994; Braun 1994; Braun and Evans 1991,1995; Burnside, Eichenbaum, and Rebelo 1993; Cecchetti, Lam, and Mark 1993; Christiano and Eichenbaum 1992; den Haan 1995; Fisher 1993; Marshall 1992; and Reynolds 1992.) For the most part, the theory available for conducting inference with these tools is asymptotic. Recently, efforts have been made to investigate the finite sample accuracy of the asymptotic theory. Much of this work has focused on the sampling distribution of statistics used in the empirical analysis of asset pricing and inventory investment models. Analyses in the context of studies of asset pricing theories include Burnside (1991); Ferson and Foerster (1991); Kocherlakota (1990); Neely (1993); and Tauchen (1986). West and Wilcox (1992) and Fuhrer, Moore, and Schuh (1993) conduct finite sample studies of inference in the context of inventory investment models. Analyses of the finite sample properties of instrumental variables estimation include Christiano (1989), Ericsson (1991), and Nelson and Startz (1990).

This paper uses Monte Carlo methods to investigate the finite sample properties of statistics often used in the analysis of business cycles. We are particularly interested in the finite sample performance of GMM for conducting inference about correlations, standard deviations, and relative standard deviations of data that have been filtered to induce covariance stationarity. We focus on the two filters most often used in business cycle analysis: the first difference filter and the Hodrick-Prescott (HP) filter. Statistics based on these filters provide different information about the data because the filters emphasize different frequencies. We also examine the finite sample properties of the GMM procedures implemented in Christiano and Eichenbaum (1992) for testing the null hypothesis that a model's implications for a set of second moments correspond with the actual second moment properties of the data.

To calculate the asymptotic standard errors using GMM, one needs to estimate the zero-frequency spectral density of a particular disturbance process. We estimate this using the heteroskedasticity and autocorrelation consistent covariance matrix estimators (HAC) described in Newey and West (1987,1994), Andrews (1991), and Andrews and Monahan (1992). The estimators differ according to the choice of kernel, the bandwidth parameter, and the order of prewhitening. One factor that distinguishes our investigation from previous ones is our analysis of the HP filter. Our statistical environment (chosen for empirical plausibility) has the property that the HP filter introduces a complicated serial correlation pattern into the relevant GMM disturbance process. As is well known, persistence puts zero-frequency spectral density estimators to a severe test.

We begin our analysis by investigating the coverage probabilities of confidence intervals computed for various second moments. We first discuss these issues thoroughly in the context of a univariate data generating mechanism (dgm) that has proved useful in the analysis of several macroeconomic data series. The advantage of this dgm, aside from its empirical plausibility, is that its simplicity enables one to develop intuition about the reasons that

alternative HAC estimators have different finite sample properties. We then analyze the finite sample properties of these HAC estimators using artificial data generated by a multivariate dgm, estimated by fitting a vector autoregression to the set of postwar U.S. macroeconomic data typically analyzed in business cycle studies.

Next, we evaluate the finite sample properties of the chi-square test implemented in Christiano and Eichenbaum (1992) for testing the fit of an equilibrium business cycle model. The test compares a model's implications for second moments with the actual second moments estimated in the data. The test takes into account the joint sampling uncertainty in model parameter estimates and data second moments. Our Monte Carlo study is based on data generated from the Long and Plosser (1982) model, the solution of which may be obtained analytically.

The paper is organized as follows. The next section lays out the asymptotic sampling theory that is relevant to our analysis. Section 3 discusses the different procedures to estimate the spectral density at frequency zero. Section 4 describes the features of the HP filter that are relevant to our analysis. Section 5 presents the analysis of inference about second moment properties. Section 6 analyzes the finite sample properties of the test of an equilibrium model. Finally, section 7 presents a summary of our main findings and concludes.

2 Generalized Method of Moments

In this section we discuss the use of GMM for the estimation of parameters and for testing hypotheses. (For textbook treatments, see Davidson and MacKinnon 1993, Hamilton 1994, and Ogaki 1992.) In the first subsection, we survey the relevant large sample theory. In the second subsection we discuss hypothesis testing.

2.1 Large Sample Theory

Suppose we wish to estimate an $n \times 1$ vector, ψ^o, of parameters. To do so using GMM, we need to first identify an $n \times 1$ vector, $u_t(\psi)$, which is a strictly stationary stochastic process for each $\psi \epsilon \Re^n$, and which has the property

$$Eu_t(\psi^o) = 0, \tag{1}$$

where ψ^o denotes the true values of the parameters of the underlying data generating mechanism. (For other regularity conditions on $u_t(\psi)$, see Hansen (1982).) We consider the exactly identified case only, so that the dimensions of the GMM error, u_t, and the parameter vector, ψ, coincide. For an analysis of finite sample issues in the over-identified case, see Burnside and Eichenbaum (1994). In the exactly identified case, the GMM estimator of ψ^o, denoted $\widehat{\psi}_T$, is defined by

$$g_T(\widehat{\psi}_T) = 0, g_T(\psi) \equiv \frac{1}{T}\sum_{t=1}^{T} u_t(\psi), \tag{2}$$

where T denotes the sample size. According to Hansen (1982), $\widehat{\psi}_T$ has the following asymptotic distribution:

$$\sqrt{T}(\widehat{\psi}_T - \psi^o) \sim N(0, V), \tag{3}$$

where

$$V = D^{-1}S(D')^{-1}. \tag{4}$$

Here, D is given by

$$D = E[\frac{\partial g_T(\psi)}{\partial \psi'} \mid_{\psi=\psi^o}], \tag{5}$$

and $'$ denotes transposition. Also, S is the (positive definite) spectral density at frequency zero of $u_t(\psi^o)$, defined by

$$S = \sum_{l=-\infty}^{\infty} C_l \tag{6}$$

and

$$C_l = E\ u_t(\psi)\ u_{t-l}(\psi)'. \tag{7}$$

Since V is unknown, in practice it must be replaced by a consistent sample estimate, $\widehat{V}_T$, which is based on replacing D and S in (4) by $\widehat{D}_T$ and $\widehat{S}_T$. Here, $\widehat{D}_T$ is computed by replacing the expectation operator in (5) by the sample average operator and ψ^o by $\widehat{\psi}_T$. Also, $\widehat{S}_T$ is obtained by applying an estimator of the spectral density at frequency zero to $u_t(\widehat{\psi}_T)$. Thus, in practice, inference is conducted based on

$$\widehat{V}_T = \widehat{D}_T^{-1}\widehat{S}_T(\widehat{D}_T')^{-1}. \tag{8}$$

We discuss alternative estimators, $\widehat{S}_T$, in section 3.

2.2 Hypothesis Testing

To test a hypothesis about the i^{th} element of ψ^o, ψ_i^o, we can make use of the fact that, asymptotically,

$$\frac{\widehat{\psi}_{i,T} - \psi_i^o}{\sqrt{(\widehat{V}_{ii,T}/T)}} \sim N(0,1), \tag{9}$$

where $\widehat{\psi}_{i,T}$ is the i^{th} element of $\widehat{\psi}_T$ and $\widehat{V}_{ii,T}$ is the i^{th} diagonal element of $\widehat{V}_T$.

We will also consider tests of joint hypotheses about the elements of ψ^o. Let F be a differentiable function which maps $\Re^n$ into the $m \times 1$ zero vector, 0_m. Then $F(\psi^o) = 0_m$ represents m hypotheses, each of which potentially involves all elements of ψ^o. To test the null hypothesis, $F(\psi^o) = 0_m$, we make use of the fact that, if indeed $F(\psi^o) = 0_m$, then asymptotically,

$$\sqrt{T}F(\widehat{\psi}_T) \sim N(0_m, V_F), V_F = f(\psi^o)Vf(\psi^o)', \tag{10}$$

where the $m \times m$ matrix $f(\psi^o)$ is defined as follows:

$$f(\psi^o) = \frac{\partial F}{\partial \psi'}_{|\psi=\psi^o}.$$

In practice, V_F in (10), which depends on unknown parameters, is estimated by replacing ψ^o with $\widehat{\psi}_T$ and V with $\widehat{V}_T$:

$$\widehat{V}_{F,T} = f(\widehat{\psi}_T)\widehat{V}_T f(\widehat{\psi}_T)'. \tag{11}$$

We base inference on the asymptotic result:

$$TF(\widehat{\psi}_T)'[\widehat{V}_{F,T}]^{-1}F(\widehat{\psi}_T) \sim \chi_m. \tag{12}$$

The popular 'calibration' procedure of Kydland and Prescott (1982) tests a restriction of the form $F(\psi^o) = 0_m$. The procedure calculates the second moments implied by an economic model at the estimated values of its parameters and compares these second moments with the second moments observed in the data. Equation (12) constitutes a formal theory of inference that can potentially be of assistance in making this comparison. It takes into account the joint sampling uncertainty in the model parameter estimates and data second moments.

3 Estimation of a Spectral Density at Frequency Zero

This section discusses the computation of $\hat{S}_T$, an estimator of the spectral density of $u_t(\psi^0)$ at frequency zero. This object is central to econometric analyses using GMM. In section 2 we showed that in the exactly identified case, $\hat{S}_T$ is required for conducting hypothesis tests on the parameters, ψ. In the over-identified case not considered in this paper, $\hat{S}_T$ also plays a role in computing the point estimates, $\hat{\psi}_T$. (See Hansen 1982.) In our simulation analysis we study five zero-frequency spectral density estimators, $\hat{S}_T$, and these are described below.

It is useful for us to describe our estimators of S by reference to the following baseline class of nonparametric estimators. (See den Haan and Levin 1994 for an analysis of parametric estimators.)

$$\widehat{S}_T = \sum_{j=-T+1}^{T-1} \kappa(j)\widehat{C}_j, \tag{13}$$

where $\kappa(\cdot)$ is a weighting function (kernel) to be discussed below and

$$\widehat{C}_l = \frac{1}{T-n}\sum_{t=l+1}^{T} u_t(\widehat{\psi}_T)\, u_{t-l}(\widehat{\psi}_T)', l = 0, ..., T-1, \tag{14}$$

where

$$\widehat{C}_l = \widehat{C}'_{-l}, l = -1, -2, ..., -T+1.$$

In (14), the scalar, n, is included as a finite sample correction. In the first subsection below, we discuss a perturbation on the above estimator that was described by Andrews and Monahan (1992) and involves 'prewhitening' $u_t(\hat{\psi}_T)$. In the second two subsections, we discuss aspects of the problem of choosing the kernel, κ.

3.1 Prewhitening

Andrews and Monahan (1992) propose and study a modification to the class of estimators defined by (13), which involves prewhitening $u_t(\widehat{\psi}_T)$. Their procedure is executed in three steps. In the prewhitening step, compute $\hat{u}_t^*(\widehat{\psi}_T)$, the fitted residual from the following b^{th} order vector autoregression to $u_t(\widehat{\psi}_T)$:

$$u_t(\widehat{\psi}_T) = \sum_{r=1}^{b} \widehat{A}_r u_{t-r}(\widehat{\psi}_T) + \hat{u}_t^*(\widehat{\psi}_T), \tag{15}$$

where $\hat{u}_t^*(\widehat{\psi}_T) \equiv u_t(\widehat{\psi}_T)$ when $b = 0$. The second step applies an estimator of the spectral density at frequency zero to $\hat{u}_t^*(\widehat{\psi}_T)$:

$$\widehat{S}_T^* = \sum_{j=-T+1}^{T-1} \kappa(j)\widehat{C}_j^*, \tag{16}$$

where $\widehat{C}_j^*$ is the j^{th} autocovariance of $\hat{u}_t^*(\widehat{\psi}_T)$, computed using the analog of (14), and $\kappa(\cdot)$ is a real-valued kernel discussed below. Finally, the prewhitened estimator of S, $\widehat{S}_T^b$, is

$$\widehat{S}_T^b = (I - \sum_{r=1}^{b} \widehat{A}_r)^{-1} \widehat{S}_T^* [(I - \sum_{r=1}^{b} \widehat{A}_r)']^{-1}. \tag{17}$$

We define $\widehat{S}_T^b \equiv \widehat{S}_T^* = \widehat{S}_T$ when $b = 0$, where $\widehat{S}_T$ is defined in (13). Andrews and Monahan (1992) give no advice on the appropriate choice of b. In their Monte Carlo studies, they consider $b = 0$ and $b = 1$, and we consider $b = 0, 1, 2$.

3.2 Alternative Choices of the Kernel, κ

We now discuss the choice of the kernel, $\kappa(\cdot)$. Newey and West (1987) use the Bartlett kernel, that is,

$$\kappa(j) = [1 - \frac{\mid j \mid}{\xi}]^{\theta}, \ 0 \leq |j/\xi| \leq 1, \ \kappa(j) = 0, \mid j/\xi \mid > 1, \tag{18}$$

with $\theta = 1$. We refer to (17) with the Bartlett kernel as the *Bartlett estimator of S, with bandwidth ξ and prewhitening parameter b*. An alternative kernel, proposed by Hansen and Hodrick (1980) and White (1984), is (18) with $\theta = 0$. We refer to this as the unweighted, truncated kernel. An advantage of the Bartlett kernel is that positive definiteness of $\widehat{S}_T$ is guaranteed, while this is not the case for the unweighted, truncated kernel. To accommodate the latter observation, we define the *unweighted, truncated estimator of S, with bandwidth ξ and prewhitening parameter b*, as one which uses the truncated kernel to compute $\hat{S}_T^*$ in (17) when $\widehat{S}_T^b$ is positive definite and the Bartlett estimator otherwise. In the data generating processes considered in this paper, we found that failure of positive definiteness occurs with low probability. In particular, in all of our experiments involving artificial data sets of length

120 observations, the frequency with which positive definiteness fails never exceeds 6% and is typically closer to 1%. We also considered data sets of length 1000, and in these we never encountered the problem.

We also consider the QS kernel proposed in Andrews (1991):

$$\kappa(j) = \kappa_{QS}(j/\xi), \tag{19}$$

where

$$\kappa_{QS}(x) = \frac{25}{12\pi^2 x^2}\left[\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right] \tag{20}$$

with $\kappa_{QS}(0) \equiv 1$. Like the Bartlett kernel, this kernel guarantees a positive definite estimator of S.

3.3 Automatic Bandwidth Selection

In this subsection, we describe data-based ('automatic') methods that select bandwidths for the Bartlett and QS kernels. We denote a bandwidth that is selected as a function of the data by ξ_T. Andrews (1991) and Newey and West (1994) (Newey-West) each describe methods that can be used to compute ξ_T for the Bartlett, QS, and other kernels. Although when $n > 1$, their efficiency criterion guiding the selection of ξ_T differs slightly, the primary difference between Andrews and Newey-West lies in their strategies for exploiting the information in the sample autocorrelation function to select a value for ξ_T. The Andrews procedure, as implemented by Andrews and Monahan, assumes an AR(1) parametric structure for the autocovariance function, which enables the analyst to select the truncation parameter based on just the variance and first order autocovariance. By contrast, Newey and West do not assume a parametric structure, and so their procedure must select the lag length based on a longer list of autocorrelations.

Neither method is entirely automatic in the sense that it completely avoids selecting parameters exogenously. In the case of the Andrews method, a time series model must be selected for $\hat{u}_t^*(\hat{\psi}_T)$. No automatic procedure is offered for doing this, though Andrews and Monahan (1992) do recommend (on computational tractability grounds, it seems) the AR(1) model. Similarly, the Newey-West method requires picking a bandwidth parameter (not ξ itself!) exogenously, and in their work they use two arbitrarily chosen values for it.

For the kernels discussed in the previous subsection, consistency of $\widehat{S}_T^*$, i.e., $p\lim \widehat{S}_T^* = S^*$, is guaranteed if $\xi_T \to \infty$ as $T \to \infty$, with $\xi_T/T^{1/2} \to 0$. (See Andrews (1991).) Andrews and Newey-West select $\{\xi_T\}$ from this class to optimize asymptotic efficiency criteria. The optimal choice of ξ_T is

$$\xi_T = 1.1447[\alpha(1)T]^{1/3} \tag{21}$$

for the Bartlett kernel and

$$\xi_T = 1.3221[\alpha(2)T]^{1/5} \tag{22}$$

for the QS kernel. (See Andrews and Newey-West and the references they cite.) In practice, in (21)-(22), $\alpha(1)$ and $\alpha(2)$ must be replaced by sample estimates.

Under regularity conditions satisfied here, Andrews and Newey and West show that (21)-(22) with the α's replaced by particular sample estimates also optimize their respective efficiency criteria when the underlying parameters, ψ, are unknown, and $b = 0$. (See Andrews and Monahan (1992) for the $b > 0$ case, where $A_1, \ldots, A_b$ must be estimated too.) We turn now to a description of the details of the Andrews optimal bandwidth selection procedure and the Newey-West optimal bandwidth selection procedure.

The Andrews Bandwidth Selection Procedure

Andrews proposes estimating the parameters of a time series model for $\hat{u}_t^*(\hat{\psi}_T)$ and provides formulas for estimating $\alpha(1)$ and $\alpha(2)$ based on the parameter estimates. In practice, Andrews and Monahan recommend fitting an AR(1) representation to the a^{th} component of $\hat{u}_t^*(\hat{\psi}_T)$, $a = 1, ..., n$. Letting $(\widehat{\rho}_a, \widehat{\sigma}_a^2)$ denote the associated first order autoregressive and innovation variance parameters,

$$\hat{\alpha}(1) = \frac{\sum_{a=1}^{n} \omega_a \frac{4\widehat{\rho}_a^2 \widehat{\sigma}_a^4}{(1-\widehat{\rho}_a)^6(1+\widehat{\rho}_a)^2}}{\sum_{a=1}^{n} \omega_a \frac{\widehat{\sigma}_a^4}{(1-\widehat{\rho}_a)^4}} \tag{23}$$

and

$$\hat{\alpha}(2) = \frac{\sum_{a=1}^{n} \omega_a \frac{4\widehat{\rho}_a^2 \widehat{\sigma}_a^4}{(1-\widehat{\rho}_a)^8}}{\sum_{a=1}^{n} \omega_a \frac{\widehat{\sigma}_a^4}{(1-\widehat{\rho}_a)^4}}. \tag{24}$$

We follow Andrews and Monahan, who suggest setting $\omega_a = 1$ for all a.

To summarize, the Andrews bandwidth selection procedure is implemented in the following four steps:

1. Obtain a series, $u_t(\hat{\psi}_T)$, and compute the fitted residuals, $\hat{u}_t^*(\hat{\psi}_T)$, in (15) if $b \geq 1$. If $b = 0$, then $\hat{u}_t^*(\hat{\psi}_T) = u_t(\hat{\psi}_T)$.

2. Fit scalar AR(1) representations to each of the n elements of $\hat{u}_t^*(\hat{\psi}_T)$. Denote the resulting parameter estimates by $(\widehat{\rho}_a, \widehat{\sigma}_a^2)$, $a = 1, \ldots, n$.

3. Select a set of weights, ω_a, $a = 1, \ldots, n$, and compute $\hat{\alpha}(1), \hat{\alpha}(2)$ using (23)-(24).

4. Evaluate (21)-(22) with $\alpha(q)$ replaced by $\hat{\alpha}(q)$, $q = 1, 2$.

We refer to the procedure for computing $\hat{S}_T^*$ which uses the QS kernel and the above bandwidth selection method as the *Andrews (QS) estimator of S, with prewhitening parameter b*. We refer to the procedure which uses the Bartlett kernel and the above bandwidth selection method as the *Andrews (Bartlett) estimator of S, with prewhitening parameter b*.

The Newey-West Bandwidth Selection Procedure

Newey and West's formulas for $\hat{\alpha}(1)$ and $\hat{\alpha}(2)$ are as follows:

$$\hat{\alpha}(q) = \left[\frac{w'\hat{F}^{(q)}w}{w'\hat{F}^{(0)}w}\right]^2, \; q = 1, 2, \tag{25}$$

and they make w a vector of 1's. Here,

$$\hat{F}^{(q)} = \sum_{j=-l}^{l} |j|^q \hat{C}_j^*, l = \beta(T/100)^{2/9}. \tag{26}$$

Newey and West avoid making parametric assumptions about the time series representation of $\hat{u}_t(\hat{\psi}_T)$. However, they must choose an exogenous bandwidth parameter, β. They suggest doing so by trying alternative values and 'then exercising some judgment about sensitivity of results' (p.7). In practice, they work with values of β equal to 4 and 12. In our Monte Carlo experiments we used $\beta = 4$. We found very little differences between setting β equal to 4 or 9. To summarize, the Newey-West bandwidth selection procedure is implemented in the following four steps:

1. Same as step 1 in the Andrews procedure.
2. Set β and compute $\hat{F}^{(q)}$, $q = 0, 1, 2$, using (26).
3. Select a set of weights, w, and compute $\hat{\alpha}(q)$, $q = 1, 2$, using (25).
4. Evaluate (21)-(22) with $\alpha(q)$ replaced by $\hat{\alpha}(q)$, $q = 1, 2$, and retain the integer value of ξ_T.

We refer to the procedure for computing $\hat{S}_T^*$ which uses the Bartlett kernel and the above bandwidth selection method as the *Newey-West (Bartlett) estimator of S, with prewhitening parameter b*. For convenience, the various zero-frequency spectral estimators and their names are summarized in Table 1.

4 The Hodrick-Prescott Filter

We consider two detrending methods: first differencing and applying the HP filter. We briefly review properties of the HP filter which are relevant to our analysis.

Suppose we have a partial realization of length T, $Y = [Y_{-T/2+1}, \ldots, Y_{T/2}]'$, of a stochastic process, $\{Y_t\}$. Application of the HP filter to this partial realization first involves computing a T dimensional trend path, $\tau = [\tau_{-T/2+1}, \ldots, \tau_{T/2}]'$, which minimizes

$$\sum_{t=-T/2+1}^{T/2} (Y_t - \tau_t)^2 + \lambda \sum_{t=-T/2+2}^{T/2-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2, \tag{27}$$

with λ normally being set to 1600 with quarterly data. The detrended series is $Y^d = [Y^d_{-T/2+1}, \ldots, Y^d_{T/2}]'$, where $Y^d_t = Y_t - \tau_t$. As pointed out in Prescott (1986), the solution to this problem can be represented as follows:

$$Y^d = A_T Y, \tag{28}$$

where A_T is a $T \times T$ matrix with elements that depend upon the values of λ and T, but not upon the data, Y. Thus, the weights in the HP filter, the rows of A_T, are a function of T. The weights are graphed in Figure 1a for $T = 120$ and $\lambda = 1600$. The figure displays the entries in rows 2, 10, 25, 60, 105, 110, and 118 of A_{120}. These are the filter weights for Y^d_t, $t = -58, -50, -35, 0, 35, 50$, and 58, respectively. The filter weights used to get Y^d_0 are essentially the HP filter weights for $T = \infty$. The figure indicates that these weights extend forward and backward in time a little over 25 periods. For this reason, the 25^{th} and 95^{th} rows of A_{120} show some (slight) evidence of truncation. For observations on Y^d_t that are less than 25 periods from the endpoints of the data set, the HP filter weights are significantly different from their $T = \infty$ values.

There is a simple representation of the HP filter weights for large T. As shown by King and Rebelo (1993), manipulation of the first order conditions of the above optimization problem shows that, as $T \to \infty$, $Y^d_t \to g(L)Y_t$, where, for $\lambda = 1600$,

$$g(L) = 0.7794\frac{(1-L^{-1})^2(1-L)^2}{h(L)h(L^{-1})}, \quad h(L) = 1 - 1.7771L + 0.7994L. \tag{29}$$

This result provides the sense in which, for large T, the HP filter induces covariance stationarity in processes that require up to fourth differencing to induce stationarity. (Examples include processes that are integrated of order up to four.) In addition, the preceding discussion suggests that for $\lambda = 1600$ and finite $T > 40$, HP filtered observations in the middle of a data set virtually coincide with $g(L)Y_t$, a covariance stationary process. These considerations have lead researchers to conclude that the HP filter, as conventionally applied, is equivalent to the application of $g(L)$ to the data, together with a particular strategy for dealing with endpoints. Baxter and King (1994) also discuss the endpoint issue, and they suggest dealing with it by dropping observations at the beginning and the end of the data set. Although it is beyond the scope of this paper to do so, it would be of interest to compare the sampling properties of this strategy for dealing with endpoints with the conventional strategy.

In the introduction we drew attention to the fact that the HP and first difference filters emphasize different frequencies. (See also Singleton 1988.) This is illustrated in Figure 1b, which shows how the first difference and $g(L)$ filters scale the spectrum of a raw time series. The horizontal axis reports ω, which is frequencies divided by π, and the vertical axis reports the transfer function of the filter. Cycle periods are given by $2/\omega$. Since business cycle analysis is primarily concerned with quarterly data, we think of the period as being one quarter. As is well known, the first difference filter amplifies the high (quarters 8 and lower) frequency components of the data, while reducing the lower frequency components. The HP filter resembles the high-pass filter also displayed in the figure, which eliminates

cycles of period 32 quarters and greater and leaves shorter cycles untouched. The figure reveals why some business cycle analysts prefer the HP filter. To them, using first difference adjusted data to study business cycles is much like using seasonally adjusted data to study the seasonal cycle. Business cycle frequencies are commonly associated with periods 8 through 32 quarters (i.e., $\omega = 0.063$ to 0.25), and the first difference filter dramatically reduces the relative importance of these.

At the same time, there are reasons that some researchers prefer to work with the first difference filter. For example, for some researchers the variable of interest may be *defined* in terms of the first difference filter because, say, they are interested in the growth rate of GDP, or inflation, rather than the levels of these variables. In this paper, we simply take it as given that, for a variety of reasons, some researchers work with the HP filter and others with the first difference filter. Our task here is to provide information on the small sample distribution of various statistics computed based on these two filters.

5 Inference About Second Moments

Our Monte Carlo experimental design is motivated by a desire to provide evidence on the finite sample properties of statistics commonly used in the analysis of business cycles. Therefore, it is important to us that (i) we study statistics that are actually in use, and (ii) we employ empirically plausible data generating mechanisms for our experiments. The first subsection below uses a univariate data generating mechanism often used in the analysis of macroeconomic data. An advantage of using this model is that its simplicity allows us to gain intuition into the basic results. In the context of this data generating mechanism, we study the finite sample performance of a standard deviation estimator. We then proceed to analyze a multivariate data generating mechanism, obtained by fitting a four variable vector autoregression to the main macroeconomic data series: consumption, employment, output, and investment. Here, we analyze the finite sample properties of 23 second moments commonly studied in the analysis of business cycles. The insights obtained in the univariate environment provide intuition into the qualitatively similar results that we get in our multivariate setting.

5.1 A Univariate DGM

We suppose the data, x_t, are generated by

$$x_t = x_{t-1} + \nu_t, \tag{30}$$

where

$$\nu_t = \rho\nu_{t-1} + \sigma\varepsilon_t,\ \varepsilon_t \sim N(0,1), t = -T/2+1, ..., T/2,$$

$|\ \rho\ | < 1$ and $\sigma > 0$. This data generating mechanism has two advantages. First, its simplicity is helpful for diagnosing our results. Second, it is a plausible representation for several U.S.

time series. (Christiano 1992 argues that this is a good representation for log GNP. Cooley and Hansen 1989 and den Haan 1995 use this to model money growth.)

We consider the problem of estimating the standard deviation of detrended x_t, x_t^d. Consequently, our parameter vector ψ is composed of a single element (i.e., $\psi^o = [E(x_t^d)^2]^{1/2}$, and $n = 1$.) It is readily confirmed that, when x_t^d is obtained by first differencing, the following specification of $u_t(\psi)$ satisfies the conditions discussed in the previous section:

$$u_t(\psi) = (x_t^d)^2 - (\psi)^2. \tag{31}$$

Note that we commit a slight abuse of notation here, because the value of ψ depends of course on the method of detrending. The value of ψ^o is unambiguous for the case of first differenced data. However, this is not so when data have been HP filtered. We confront this first, before reporting the results of our Monte Carlo experiments.

5.1.1 The Standard Deviation of HP Filtered Data

When x_t^d is obtained by HP filtering x_t, (31) does not exactly satisfy the conditions in sections 2 and 3. Endpoint effects associated with the filter have the implication that x_t^d and, hence, $u_t(\psi)$ are not strictly stationary. To quantify this, we computed $\psi_t^0 = [E(x_t^d)^2]^{1/2}$ for $t = -T/2+1, ..., T/2$ and $T = 120$. For comparison, we also computed $\tilde{\psi}_t^0 = [E(\tilde{x}_t^d)^2]^{1/2}$ for $t = -T/2+1, ..., T/2$ and $T = 120$, where $\tilde{x}_t^d = g(L)x_t$. We refer to $g(L)$ as the *infeasible HP filter*, since it requires having an infinite number (actually, 25 or so will do) of data points prior to the first observation and after the last observation in the data set. By contrast, the *feasible HP filter* requires only the available data. We computed ψ_t^0 and $\tilde{\psi}_t^0$ using a Monte Carlo analysis with 100,000 replications, in which $\rho = 0.4$ and $\sigma = 0.01$. For the experiments with the feasible HP filter, each replication is composed of 220 observations, with the starting value of x_t set to zero and the first 100 observations discarded in order to randomize initial conditions. The feasible HP filter was then computed using the remaining 120 observations. For the experiments related to the infeasible HP filter, 600 observations were generated per replication, with the first 100 deleted from the analysis to randomize initial conditions. To approximate the infeasible HP filter, we applied the feasible HP filter with A_{500} to the remaining observations and then kept the middle 120 observations for analysis. These calculations produced $\tilde{x}_{t,i}^d$ and $x_{t,i}^d$ for $t = -59, ..., 60$, and $i = 1, ..., 100,000$:

$$\psi_t^0 = [\frac{1}{100,000} \sum_{i=1}^{100,000} (x_{t,i}^d)^2]^{\frac{1}{2}}, \ \tilde{\psi}_t^0 = [\frac{1}{100,000} \sum_{i=1}^{100,000} (\tilde{x}_{t,i}^d)^2]^{\frac{1}{2}},$$

for $t = -59, ..., 60$.

These objects are graphed in Figure 1c. In addition, we display $Plim_{T\to\infty} \hat{\psi}_T$, which is 0.01877 and was computed by inverse-Fourier transforming the spectrum of $g(L)x_t$. Note the substantial variation in ψ_t^0 at the beginning and at the end of the data set, revealing that $(x_t^d)^2$ is not stationary in the mean. (See Baxter and King 1994 for a similar result using a different data generating mechanism.) By contrast, in the middle 60 observations ψ_t^0

is roughly constant and equal to $\tilde{\psi}_t^0$, which in turn essentially coincides with $Plim_{T\to\infty}\ \hat{\psi}_T$. These findings are consistent with the observations about the nature of the HP filter made in the previous section.

The endpoint effects in x_t^d present a problem for us. The asymptotic theory requires that ψ^0 in the numerator of (9) be the standard deviation of x_t^d. But there is no such number, independent of t. There are at least three options. The first is to equate in ψ^0 with $\psi_t^0 = [E(x_t^d)^2]^{1/2}$ for observations in the middle of an HP filtered data set. A feature of this option is that it overstates the degree of variation in x_t^d. For example, the mean value of ψ_t^0 for all values of t is 0.01813. A second option is to equate ψ^0 with $E\hat{\psi}_T$, which is 0.01793 for $T = 120$. (This was approximated by averaging over 1000 Monte Carlo replications of $\hat{\psi}_T$.) The discussion in the previous paragraph suggests that these two options are asymptotically equivalent. For example, when $T = 1000$, $E\hat{\psi}_T = 0.01864$, which is nearly equivalent to $Plim_{T\to\infty}\ \hat{\psi}_T$. The third option is to throw away the first and last 25 observations. We did not implement this option because we are interested in implementing the HP filter as it is used in practice. We decided to go with the second option for two reasons. First, this way of selecting ψ^0 seems closest in spirit to equating ψ^0 with the 'true' standard deviation of x_t^d. Second, this choice is conservative from the point of view of the conclusions of our analysis. We will show that asymptotic theory does not work well in finite samples. Evidence presented below suggests that, had we pursued the first option, it would have worked even worse. For the sake of comparability, we treat ψ^0 in the case of the first difference filter analogously.

5.1.2 Results Based on No Prewhitening

Our results for the case $b = 0$ can be summarized as follows. We show that there is substantial distortion — in terms of fat tails and skewness — in estimated confidence intervals for the standard deviation of x_t^d. This reflects two features of the sampling distribution of $\hat{V}_T$: it is biased downward and covaries positively with $\hat{\psi}_T$. The former is the principal reason for the fat tails, while the latter accounts for skewness. The downward bias in $\hat{V}_T$ principally reflects the persistence in $u_t(\psi)$, particularly when the data have been HP filtered. Persistence leads to distortions in part by requiring a large lag bandwidth, which the methods we implement tend to underestimate, accounting in part for the downward bias in $\hat{V}_T$. Finally, the performance of our various estimators is very similar.

These results reflect six observations, based principally on an examination of the $b = 0$, $\rho = 0.4$, $\sigma = 0.01$ case. For money and GNP growth, this value of ρ is the empirically relevant one. First, for both the HP and first difference filters, there is a substantial amount of skewness in the t-statistic defined in equation (9) when $b = 0$ and the number of observations, T, is 120. (See Table 2a, top panel.) For each variance estimator, the frequency of the associated t-statistic being in the nominal lower 5% tail is around 19% in the case of the HP filter and around 10% in the first difference case. We examined the impact on our results of identifying ψ^0 with $Plim\hat{\psi}_T$ rather than with $E\hat{\psi}_T$, and consistent with the discussion in section 5.1.1, we found that this exacerbates the skewness problem. For example, the

BART(11) row in the top left panel of Table 2a becomes 26.0, 33.4, 9.3, and 6.0. Second, when $T = 120$ there is also a substantial fat-tail problem in that the sum of the probabilities of being in the top and bottom 5% tails substantially exceeds 10%, and similarly for the top and bottom 10% tails. We also investigated the consequences of applying the infeasible HP filter and found that our results regarding skewness and fat tails are essentially unaffected by this change. This suggests that these problems do not reflect the endpoint features of the HP filter.

Third, the finite sample distortions appear to reflect problems with $\widehat{V}_T$. This can be seen by noting that there is almost no fat-tail or skewness problem in the row corresponding to TRUE. This suggests that the distribution of the numerator in (9) is nearly normal and indicates that the skewness and fat-tail problems arise almost entirely from the sampling properties of $\widehat{V}_T$. Results in column I of Table 3a bear out this view, as it applies to skewness. They show that the numerator and denominator of (9) are positively correlated. Thus, in replications when the numerator in (9) is big, the ratio is not big since the denominator is typically large too. Similarly, the impact on the ratio of a negative, but large in absolute value, realization in the numerator is typically magnified by a small denominator. Results in columns II and III of Table 3a identify problems in the sampling distribution of $\widehat{V}_T$ which may account for the fat-tail problem evident in Table 2. In particular, they show that when $b = 0$, the GMM estimator, $\widehat{V}_T$, is biased downward.. Other things the same, this would be expected to blow up tail areas.

Fourth, distortions appear to reflect the persistence in $u_t(\psi)$. The literature on small sample properties of variance estimators notes that high temporal dependence can lead to coverage probabilities for confidence intervals that are too low, i.e., that lead to excessive rejections. (See Andrews 1991, Andrews and Monahan 1992, and Ericsson 1991.) The impact of persistence on our results can be seen in two ways: by comparing results based on the HP and first difference filters and by comparing results based on $\rho = 0.4$ and $\rho = 0.1$.

Distortions — in both long and short samples — appear to be substantially lower for computations based on the first difference filter than for computations based on the HP filter. And the HP filter leaves substantially more temporal dependence in x_t^d than does the first difference filter. One way to see this can be seen in the results in Figure 1d. That figure displays four autocorrelation functions for $u_t(\hat{\psi}_T)$, where $u_t(\psi)$ is defined in (31). Each autocorrelation function is based on 100,000 observations of artificial data generated from (30) and is differentiated according to whether the data have been HP filtered or first differenced and whether $\rho = 0$ or $\rho = 0.4$. The lowest two autocorrelation functions are based on first difference filtering. The higher two are based on HP filtering. From our perspective, the notable feature of this graph is how high and relatively insensitive to ρ the autocorrelation of $u_t(\hat{\psi}_T)$ is when the underlying data have been HP filtered. (For related observations, see Cogley and Nason 1995.)

Another way to see that persistence accounts for the distortions we find is to compare the results for $\rho = 0.1$ in Table 2b with those based on $\rho = 0.4$ in Table 2a. Note that, with the drop in ρ, the coverage probabilities are somewhat closer to their nominal values in the case of first differenced data, while there is less improvement with HP filtered data. This is

consistent with the notion that persistence in x_t^d is an important factor underlying the poor small sample distribution of our test statistic. Recall from Figure 1d that a reduction in ρ substantially reduces the persistence in x_t^d when data have been first differenced, but not when data have been HP filtered.

Fifth, high persistence produces a downward bias in $\hat{V}_T$ in part because our automatic bandwidth selection procedures are themselves downward biased. To see this, we note first that the relatively high persistence in $u_t(\hat{\psi}_T)$ when data have been transformed by the HP filter implies that a high bandwidth parameter is needed to estimate the zero-frequency spectral density. Consider the results in Figures 2a and 2b. They graph $S(\xi)/S$ against values of the bandwidth parameter, $\xi = 1, ..., 101$. $S(\xi)$ is (6) with the summation truncated at $l = \xi$, using the kernel indicated in the figure. These objects were computed using a single realization of length 1000 generated from (30) with $\rho = 0.4$ and $\sigma = 0.01$. In these calculations, $b = 0$ and S is approximated by $S(101)$, computed using the unweighted, truncated kernel (i.e., (18) with $\theta = 0$.) This normalization guarantees that all the curves in Figure 2 eventually converge to unity. The curves marked 'UNWEIGHTED' correspond to the kernel in (18) with $\theta = 0$. For the curves market 'BARTLETT' we set $\theta = 1$. The curves marked 'QS' correspond to the case where the kernel is (20). Figures 2a and 2b report results based on filtering data using the HP filter and the first difference filter, respectively.

Consistent with the findings in Figure 1d, Figures 2a and 2b suggest that it takes a much higher value of ξ to get a variance estimator to converge for data based on the HP filter than for data based on the first difference filter. For example, in the case of 'BARTLETT,' the variance estimator has 90% converged for ξ less than 6 when the data have been first differenced, while a value of ξ in excess of 31 is needed to get comparable convergence when the data have been HP filtered.

The results in column I of Table 4 indicate that the automatic lag selection procedures detect the need for a higher bandwidth when data have been HP filtered. The procedures based on Andrews and Newey-West select average lag lengths of 10 and 5, respectively. The reason the former selects 10 on average reflects its 'strategy' for guessing the lag of the last significant autocorrelation: it extrapolates the first autocorrelation (0.75, in this case) and assumes $u_t(\psi)$ is AR(1). The Newey-West method with $\beta = 4$ and $T = 120$ guesses the lag of the last significant autocorrelation by 'looking' at the first four autocorrelations. It is then not surprising, based on inspection of Figure 1d, that the Newey-West method selects a lag length of 5 and misses the 'hump' in the autocorrelation function at higher lags. The results in Figure 2a suggest that, particularly for HP filtered data, the chosen bandwidths are not large enough. As the figure indicates, with too small a bandwidth, one expects the standard error estimate to be understated and tails to be blown up.

Sixth, results for the various procedures are all very similar. In this example, it makes little difference how exactly the bandwidth or kernel is picked. For example, even though the lag lengths picked by the Andrews and Newey-West methods are different, the results in Figure 2a indicate that the implied estimates of $\hat{S}$ are not very different.

Finally, the skewness and fat-tail problems are reduced when the number of observations is increased to 1000 (see the bottom half of Table 2a). This is to be expected, based on large

sample theory. But the reduction is surprisingly small, particularly for results based on the HP filter.

5.1.3 Results Based on Prewhitening

Our results in this subsection can be summarized as follows. We show that first order prewhitening has a beneficial impact on the fat-tail problem, but relatively less impact on the skewness problem. The impact on the fat-tail problem reflects that first order prewhitening reduces the downward bias in $\hat{V}_T$ that contributes to fat tails when $b = 0$. An important objective of a project such as ours is to determine which of the several existing zero spectral density estimators works best in our setting. And so we initially found it interesting that the Andrews and Andrews and Monahan zero-frequency spectral density estimator appears to outperform the Newey and West procedure when $b = 1$. However, it turns out that this result does not actually indicate any inherent superiority in the former. These procedures have a variety of potential sources of bias. In our application the results reflect that the former procedure is driven by two sources of bias which tend to cancel, while only one of these sources of bias is present in the Newey-West (NW) procedure. The two sources of bias are (i) misspecification inherent in the relatively parametric Andrews and Andrews and Monahan procedure which leads to an underestimate of the bandwidth and (ii) a bias affecting both procedures which leads to an underestimate of $S(\xi)$ for any given bandwidth, ξ. In contrast with the no prewhitening case, consideration (i) by itself leads to an overestimate of V. This is because first order prewhitening in our context induces high order negative autocorrelation in the prewhitened $u_t(\psi)$, $u_t^*(\psi)$. Finally, our results based on second order prewhitening are somewhat discouraging, since they are essentially identical to the $b = 0$ results. Moreover, we are not aware of any algorithm that would lead a researcher to select $b = 1$. We show that the AIC criterion invariably leads to a selection of $b = 2$.

These results reflect six observations, with the first five being based on $b = 1$. First, in the cases when prewhitening does help, it does so mainly by alleviating the problem of fat tails and does little to reduce the skewness problem. (Compare the $b = 1$ rows with the $b = 0$ rows in Table 2a.) For example, when $T = 120$ and data have been HP filtered, the sum of the lower and upper 5% tail area probabilities is 28.0 for QS when $b = 0$ and falls to 13.3 when $b = 1$. At the same time, the left tail area exceeds the right by 9.4% when $b = 0$ and by 10.7 when $b = 1$. The favorable impact on the fat-tail problem of increasing b is consistent with simulation results in Andrews and Monahan (1992). But they do not analyze skewness. In our example we would clearly overstate the benefits of prewhitening by abstracting from skewness.

Second, the beneficial impact on the fat-tail problem of first order prewhitening appears to reflect a rise in the mean of $\hat{V}_T$. (See column III, Table 3a and 3b.) This impact appears to have been greatest for QS and BARTLETT, and not surprisingly, these also exhibit the smallest fat-tail problem. (See Table 2a and 2b.)

Third, the differences between the automatic bandwidth selection methods are greatest with $b = 1$. For $b = 1$, NW appears to perform worst, at least relative to the fat-tail problem.

For example, the sum of the upper and lower 5% tail areas is equal to 21.8 for NW and to 13.7 and 13.3 for BARTLETT and QS, respectively. We argue that, ironically, the relatively poor performance of NW reflects that it is distorted by fewer sources of bias. To see this, we note first that selecting the bandwidth by the Newey-West procedure delivers about the same results as the much simpler procedure of simply setting ξ exogenously to 11. (Compare the BART(11) and NW rows in the HP block of columns in Tables 2a and 2b.) This can be seen in Table 4a, which shows (column I) that the mean value of ξ_T chosen by NW is roughly the same as the bandwidth in the BART(11) procedure. In contrast, the BARTLETT procedure selects a much lower ξ_T on average.

The reason that, in this example, Newey-West selects much higher ξ_T's on average than does Andrews is instructive about the differences between these procedures. When $b = 1$, it appears that the autocorrelation function of $\hat{u}_t^*$ is positive for the first few lags, after which it turns sharply negative. This can be seen in Figure 3, which is the exact analog of Figure 2a, for $b = 1$. Note the initial rise in 'UNWEIGHTED,' followed by a sharp fall, as the bandwidth increases above $\xi = 3$. To see the implications of this, recall that the two methods select the bandwidth based on different strategies for extrapolating $\hat{u}_t^*$. The AR(1) version of the Andrews procedure used here looks at the lag 0 and 1 autocovariances of $\hat{u}_t^*$ and extrapolates based on this. When $b = 1$, Figure 3 shows that this extrapolation is very misleading. The AR(1) assumption clearly entails specification error, since it completely misses the oscillatory behavior of the actual autocovariance function. Since, in addition, the first order autocorrelation is small, the Andrews procedure picks a small bandwidth. Newey-West looks at more elements in the autocorrelation function, properly detects its complexity, and therefore sets a much higher value of the bandwidth, on average.

The following simple example illustrates why, in our setting, first order prewhitening causes the Andrews and Andrews and Monahan lag length selection procedure to understate the bandwidth, while the Newey-West procedure gets it about right. Let $u_t = \varepsilon_t + \varepsilon_{t-1}$, where ε_t is uncorrelated over time. The autocorrelation function of u_t, which is relatively high at lag one ($\rho = 0.5$) and then falls to 0.0, resembles qualitatively the autocorrelation for HP filtered data exhibited in Figure 1d. The autocorrelation function of the first order prewhitened process, $u_t^* = u_t - \rho u_{t-1}$, is $\rho_1^* = 1/6$, $\rho_2^* = -1/3$ and $\rho_j^* = 0$ for $j \geq 3$ at lags 1, 2, and higher, respectively. The Andrews procedure approximates this autocorrelation function of u_t^* using an AR(1) functional form and the relatively small ρ_1^*. Substituting $\rho_1^* = 1/6$ into (23) yields $\alpha(1) = 0.12$, or, for $T = 120$, using (21), $\xi_T = 2.77$. The NW procedure, with $\beta = 4$, and $T = 120$ selects $\alpha(1) = 2.25$ using (25), or, $\xi_T = 7.40$ using (21). This example captures the reasons that the Andrews method picks a shorter lag length than Newey-West, when $b = 0$. In that case, the Andrews method implies that $\alpha(1) = 1.7778$ for $T = 120$, and, hence, $\xi_T = 6.8$. The Newey-West method implies that $\alpha(1) = 0.24$, and $\xi_T = 3.6$. It is interesting that the Newey and West method does not pick a monotonically declining bandwidth as the order of prewhitening increases.

In view of the apparently superior performance of Newey-West over Andrews in selecting the bandwidth, it is ironic that Newey-West nevertheless underperforms relative to Andrews in our Monte Carlo results. The resolution appears to lie in the effects of finite sample

bias. In particular, we have found that the mean of the finite sample analog of the curves in Figure 3 lies considerably below those curves. (That is, letting $\hat{S}_T(\xi)$ denote an estimator of $S(\xi)$, we found that $E\hat{S}_T(\xi) < S(\xi)$ for various values of ξ and various kernels.) Given the nature of the hump near the origin, methods which select a small bandwidth, in effect, overcome this small sample bias. Thus, the superior performance of the AR(1) Andrews procedure appears to reflect the offsetting effect of specification error on bias. The Newey-West procedure does relatively poorly because it also suffers from bias, but does not enjoy the compensating effects of specification error. Clearly, this result is specific to the data generating mechanism we have assumed. Still, it illustrates the kind of factors that impact on the small sample performance of alternative zero-frequency spectral density estimators.

Fourth, consider the effects of second order prewhitening, i.e., $b = 2$. Andrews and Monahan (1992) and Newey and West (1994) do not consider this case, but they give no motivation for only considering first order prewhitening. (For a further analysis, see den Haan and Levin 1994.) We had expected second order prewhitening to improve the performance of our test statistic, given the complex behavior of the autocorrelation function when $b = 1$. Also, with an AIC selection criterion the AR(2) was chosen 939 times out of 1000 data over an AR(1) to model $u_t(\hat{\psi}_T)$, for the case with $\rho = 0.4$, HP filtered data and $T = 120$. To our surprise, performance actually deteriorated and closely resembles the $b = 0$ case. Presumably, this reflects that, as in the $b = 0$ case, specification error (i.e., choosing too low a value of ξ_T - see Figure 4) and bias (for any ξ_T, $\hat{S}_T(\xi_T)$ underestimates $S(\xi)$) both work in the same direction, towards underestimating V. This accounts for the reappearance of the fat-tail problem when $b = 2$.

5.2 A Multivariate DGM

We estimated a data generating mechanism for log consumption, c_t, log GNP, y_t, log gross investment, i_t, and log hours, n_t. For this, we use the quarterly postwar U.S. data described in Christiano (1988). We impose that c_t, y_t, i_t are each integrated of order 1 and that $y_t - c_t, y_t - i_t, n_t$ are each covariance stationary. Define

$$Y_t = \begin{bmatrix} \Delta y_t \\ y_t - c_t \\ y_t - i_t \\ n_t \end{bmatrix}. \tag{32}$$

We estimated a VAR for Y_t for the period 1957:1-1984:1, and used this to simulate 500 artificial data sets, each of length 115 observations. We did this in two ways: one by a Monte Carlo procedure of drawing the disturbances from the normal distribution with variance estimated from the data (the rows marked 'N' in Tables 5a and 5b) and the other by bootstrapping the actual fitted residuals (the rows marked 'B' in Tables 5a and 5b). We implement these two procedures as a check on the robustness of our results.

In each artificial data set, we computed (9) for 23 statistics. For each statistic, we recorded the frequency of times, across data sets, that (9) was less than the nominal 5% critical value

and the frequency of times that (9) exceeded the 95% critical value. Our 23 statistics are standard in the business cycle literature. They include σ_y, σ_c/σ_y, σ_i/σ_y, σ_n/σ_y, σ_w/σ_y, σ_w/σ_n, where σ_x denotes the standard deviation of the detrended variable $x = y, c, i, n, w$ and w denotes labor productivity, that is, $w = y - n$. In addition, they include 17 correlations: $\rho_{yx}(\tau), \tau = -1, 0, 1$, for $x = c, i, n, w$; $\rho_{yy}(\tau), \tau = -1, 1$; and $\rho_{wn}(\tau), \tau = -1, 0, 1$. Our analysis was done for each of the two detrending methods. The results are similar to what we found in the univariate analysis in the previous section.

Consider Table 5a, which reports findings for σ_y. Standard error estimates, $\sqrt{\hat{V}_T}$, were based on three measures of the zero-frequency spectral density: BART(11), BARTLETT, and NW. Recall that the choice of kernel for all three procedures is the Bartlett kernel, but that the choice of bandwidth differs across these three estimators. There are three results we would like to emphasize. First, results are very similar for the experiments with Normal and with bootstrapped errors. Second, for both detrending methods there is a skewness and fat-tail problem when the data have not been prewhitened. The problem is less severe for data that have been first differenced. These findings closely resemble those reported for the analysis in the previous section. If anything, the skewness and fat-tail problem is more severe here. Third, when detrending is by HP filter, prewhitening helps the fat-tail problem a great deal, but has little impact on the skewness problem. For example, the average of the left and right tail areas is 16.6 for BART(11) when there is no prewhitening and 5.6 when there is first order prewhitening. The latter is very close to the asymptotically correct value of 5.0. At the same time, the difference between the left and right tail areas is 10.0 in each case. Prewhitening helps little when the underlying data have been first differenced. Here, too, the results closely resemble what we found in the previous subsection.

Table 5b contains a summary of the findings for the other statistics. The full set of results, available on request, are too numerous to reproduce here. In any case, the message from these results is fairly simple and corresponds closely to our findings in the previous subsection.

The left panel of Table 5b reports results based on HP filtering the data, while the right panel is based on first differencing. The columns labelled 'I' report the absolute deviation from 10% of the sum of the two tail areas, averaged over all 23 statistics. This average does not indicate the typical sign of the deviation, and so we also present columns labelled 'II,' which indicate the number of times, out of 23, that the deviation was positive. A positive deviation indicates a 'fat-tail' problem. Columns labelled 'III' report the absolute value of the difference between the left and right tail areas, averaged over the 23 statistics. This is a measure of skewness, although the absolute value operator destroys information about whether the skewness is to the left or right. Columns labelled 'IV' provide information on that, by reporting the number of times that the skewness is to the left, i.e., the number of times that the deviation is positive.

Consider columns I and II. With no prewhitening ($b = 0$) almost all statistics display a substantial fat-tail problem, which averages from about 14 to 20%, depending upon the exact statistical procedure used. The problem is considerably less severe when data have been first differenced, although it is still substantial, being on the order of from 5 to 10%. Here, as

in the example in the previous subsection, prewhitening reduces the fat-tail problem by a substantial amount when the underlying data have been HP filtered and very little when the data have been first differenced.

Columns labelled III and IV indicate that, with no prewhitening, there is a skewness problem on the order of from 4 to 7% for results based on HP filtered data. The problem is less severe when the data have been first differenced. Here, as in the previous subsection, prewhitening has little impact on the skewness problem. Column IV indicates that there is little consistency among the underlying results on the direction of skewness. In fact, some statistics do not suffer from a skewness problem.

A subset of the results are represented in Figure 5. For each statistic, for BARTLETT and NW, and for b=0, 1 we display the probability of a statistic exceeding the 95% (height of gray bar) critical value and of being less than the 5% (black bar) critical value. Results are reported for each of our 23 statistics, and the numbering code for these statistics (#1-#23) is explained in the note to the figure. Also, the normal distribution was used in simulating the dgm. The fact that there is no pattern in the direction of skewness is evident. In addition, the beneficial impact on the fat-tail problem of raising b from 0 to 1 is also evident.

Finally, whether we use the bootstrap or Normal distribution in simulating our dgm makes little difference to the results. This is consistent with the underlying asymptotic theory.

6 A Wald-Test Example

In this section we study the finite sample properties of the Wald-type statistics proposed by Christiano and Eichenbaum (1992) for testing the null hypothesis that a model's implications for the second moment properties of a set of variables coincide with the second moment properties of those variables in the data. We pursue this in a simple example.

This Wald test provides a statistic to evaluate a model's goodness of fit and overcomes an important weakness of the 'calibration' approach. The 'calibration' approach consists of two steps. In the first step the model's structural parameters, ψ_1, and some data second moments, ψ_2, are estimated. The economic model implies a relation between the structural parameters and the second moments. We denote this relation by $\psi_2 = g(\psi_1)$, or $F(\psi) = 0$. The second step consists of comparing the estimated second moments, $\widehat{\psi}_{2,T}$, with the ones implied by the economic model, $g(\widehat{\psi}_{1,T})$. Singleton (1988) points out that a disadvantage of the 'calibration' approach is that the metric of evaluating the difference between $\widehat{\psi}_{2,T}$ and $g(\widehat{\psi}_{1,T})$ is not made explicit. In section 2, we showed how standard asymptotic theory can be used to construct a formal metric of the difference between the two sets of moments.

An alternative estimation and testing strategy imposes the restriction $F(\psi) = 0$ during estimation. For example, the vector ψ could be estimated by designating ψ_1 as the free parameters and setting $\psi_2 = g(\psi_1)$. This estimation strategy would typically require the use of a nonlinear search algorithm to optimize the estimation criterion, and this would typically involve evaluating $g(\psi_1)$ hundreds, perhaps thousands, of times. A difficulty with

this strategy is that for many interesting models, it is computationally costly to calculate $g(\psi_1)$ for a particular value of ψ_1. An advantage of the Wald procedure studied here is that it involves estimating ψ without imposing the restrictions of the model. Then, for testing purposes, all that is required is the derivative of $g(\cdot)$, and numerical procedures to approximate this only require evaluating $g(\psi_1)$ a small number of times.

6.1 The Brock-Mirman Model

We use the Brock-Mirman version of the neoclassical growth model, which has log-utility and complete depreciation. As demonstrated in Long and Plosser (1982), this model has the advantage that its analytic solution is known. The finite sample properties of test statistics are, therefore, not affected by approximation error in the model solution. Den Haan and Marcet (1994) show that small numerical errors can be important for the distribution of the test statistics. More complicated examples are analyzed in Burnside (1991) and Burnside and Eichenbaum (1994).

According to the model, a planner selects contingent plans for consumption, c_t, and capital, k_{t+1}, to maximize $E_0 \sum_{t=0}^{\infty} \beta^t \log(c_t)$ subject to the resource constraint, $c_t + k_{t+1} = k_t^{\alpha} \exp(z_t)$; the exogenous technology shock process, $\Delta z_t = \rho \Delta z_{t-1} + \sigma \varepsilon_t$; and the given initial conditions, $k_0 > 0$, z_0, z_{-1}. Here, $\beta = 0.99, \sigma = 0.01$, $\alpha = 0.3$, and $\varepsilon_t \sim N(0,1)$. We consider two different values for ρ. These are $\rho = 0.1$ and $\rho = 0.5$. As is well known, the contingency plan that solves this problem is $k_{t+1} = \alpha\beta y_t$, where y_t is output and $y_t = k_t^{\alpha} \exp(z_t)$. Some simple algebra shows that $\Delta \log k_{t+1}$ is an AR(2) with parameters equal to $\alpha + \rho$ and $-\alpha\rho$. If $\rho = 0.1$, then the law of motion for $\Delta \log k_{t+1}$ is not very different from the simple example discussed in section 5.1. If $\rho = 0.5$, then the law of motion for $\Delta \log k_{t+1}$ is much more persistent.

6.2 The Wald Test

We simulated 1000 artificial data sets for this economy, of length 120, 1000, and 5000 observations each. We performed two different experiments. To simplify the Monte Carlo exercise we only estimate one model parameter, σ or ρ, and estimate one second moment, the standard deviation of detrended capital. In the first experiment we take the values of the structural parameters ρ, β, and α as given and the value of σ as unknown. In the second experiment we take the values of σ, β, and α as given and the value of ρ as unknown. We now describe the first experiment.

In each data set, we estimated a 2×1 vector, ψ, where

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma_k \end{pmatrix} \tag{33}$$

and

$$\sigma_k = [E(x_t^d)^2]^{1/2} \tag{34}$$

and x_t^d denotes detrended $\log(k_t)$. As before, x_t^d is alternatively the first difference of $\log(k_t)$ or HP filtered $\log(k_t)$. We specified the following 2×1 GMM error vector:

$$u_t(\psi) = \begin{pmatrix} (\Delta z_t - \rho \Delta z_{t-1})^2 - (\psi_1)^2 \\ (x_t^d)^2 - (\psi_2)^2 \end{pmatrix}. \tag{35}$$

It is easily established that, when detrending is accomplished by first differencing, $Eu_t(\psi^o) = 0$ and u_t satisfies the other conditions in section 2. As before, under HP filtering, $u_t(\psi)$ does not satisfy the conditions of section 2 exactly, due to the influence of endpoint effects in the application of the HP filter. However, we proceed as though the asymptotic results in section 2 are valid anyway.

Given a value of σ and of the other model parameters, it is possible to compute the model's implied variance of x_t^d. Denote this by $g(\psi_1)$. In the case of the HP filter, $g(\cdot)$ is obtained by first applying the inverse-Fourier transform to the spectral density of the HP filtered series and then taking the square root of the result. When the data are detrended by first differencing, then x_t^d is an AR(2) and we can calculate $g(\cdot)$ analytically. Then

$$F(\psi^o) = (\psi_2^o)^2 - g(\psi_1^o) = 0. \tag{36}$$

We computed $F(\widehat{\psi}_T)$ in the artificial data and compared the small sample distribution of the test statistic in (12) with the chi-square distribution with one degree of freedom. We will refer to this experiment as experiment 1. In experiment 2, the value of σ is considered known and the value ρ is estimated. Here equation (35) is replaced by

$$u_t(\psi) = \begin{pmatrix} (\Delta z_t - \psi_1 \Delta z_{t-1}) \Delta z_{t-1} - (\sigma)^2 \\ (x_t^d)^2 - (\psi_2)^2 \end{pmatrix}. \tag{37}$$

The simple nature of this example prevents us from estimating σ and ρ simultaneously and still have a meaningful Wald test. It is not hard to show that in that case the variance of $F(\widehat{\psi}_T)$ would be singular.

The results for experiments 1 and 2 are presented in Tables 6 and 7, respectively. In section 5.1, we found that results are not very sensitive to whether the Bartlett or quadratic spectral (QS) kernel is used. As a result, we only report results based on the Bartlett kernel in this section. We used the Andrews and Newey-West methods to calculate the optimal bandwidth. Consistent with the notation in section 5, here we refer to the variance estimators $(\hat{V}_T)$ based on the first and second procedure as BARTLETT and NW, respectively (see Table 1). The structure of Tables 6 and 7 is similar to that of Tables 3a and 3b in section 5. The first two columns indicate the frequency of times that the test statistic is less than the 5% and 10% critical values of the χ_1^2-distribution, respectively. The third and fourth columns indicate the frequency of times that the test statistic is greater than the 90% and 95% critical values. The fifth column contains the average value of the selected optimal bandwidth, ξ_T.

6.2.1 Experiment 1: Innovation Variance Estimated

Our results are generally consistent with the findings of section 5. Five observations are worth emphasizing. First, small sample coverage probabilities are closer to their asymptotic

values when there is little persistence, i.e., $\rho = 0.1$, than when $\rho = 0.5$. Also, as in section 5, the coverage probabilities are closer to their nominal values when data have been filtered by first differencing, rather than when they are HP filtered. In fact, the results are surprisingly good for the first difference filter, even with $T = 120$. For this reason, to save space we only report the results based on BARTLETT for the first difference filter.

Second, we found substantial skewness in the distribution of our test statistic. Moreover, this skewness primarily reflects a rightward shift relative to the chi-square distribution. This is a problem because in practice one is only interested in one-sided tests.

The third observation is that first order prewhitening increases the values of $\widehat{V}_{F,T}$, especially at low values of the bandwidth. This helps reduce the fat-tail problem. In this experiment, first order prewhitening also helps to reduce the skewness problem. Here, as in the example in section 5.1, second order prewhitening does not help. Although first order prewhitening helps, it only does so very little when data are HP filtered. The frequency of falsely rejecting the null hypothesis at the 5 or 10% levels is around twice what it should be in a sample of size 120. Fourth, as in the section 5.1 example, first order prewhitening has a bigger effect on the fat tail with BARTLETT than with NW. Fifth, as expected, the asymptotic theory is validated when the number of observations becomes large.

6.2.2 Experiment 2: Autocorrelation Estimated

The results of experiment 2, reported in Tables 7a and 7b, are similar to those of experiment 1. Here, too, NW exhibits fatter tails relative to BARTLETT when there is first order prewhitening. For example, in Table 7a ($T = 120$), the sum of the two 5% tail area probabilities is 17.5 percent for NW versus 11.3 for BARTLETT.

The reason for the poor performance of NW relative to BARTLETT when there is first order prewhitening appears to reflect that there are offsetting biases in the latter, as in the Monte Carlo analysis in section 5.1. To see this, consider first Figure 6. This figure exhibits $V_{F,70,000}$ as a function of the bandwidth used in the underlying zero-frequency spectral density calculation. Note how first order prewhitening leads to values of $V_{F,70,000}$ that are very high at low values of the bandwidth. For example, with the Bartlett kernel and first order prewhitening, a bandwidth of around 3 results in V_F being overestimated by a factor of around 2.8. In addition, in results not reported here, we find that estimates of $V_{F,T}$ are downward biased for every fixed bandwidth.

For reasons like those reported in section 5.1, the Andrews optimal bandwidth selection method computes a low value for the bandwidth (see Tables 7a and b). In a large sample, this would lead to an upward bias in $\widehat{V}_{F,T}$. However, when $T = 120$, the upward bias roughly cancels the downward bias, allowing Andrews to turn in a tolerable performance. At the same time NW, which accurately recognizes that a much larger bandwidth is needed, does poorly. Here, as in section 5.1, the problem is that NW does not have an upward bias to cancel the downward bias in estimating $\widehat{V}_{F,T}$. This is why NW produces fatter tail area probabilities.

Thus, BARTLETT's good performance relative to NW is a Pyrrhic victory. While it is

nice that BARTLETT performed well, the manner in which this was accomplished – by the offsetting effects of two biases – is cause for discomfort. There would of course be no problem if in practice the two biases in BARTLETT were perfectly correlated across applications. But that this is not the case is suggested by the results based on $T = 1000$ and $T = 5000$ in Tables 7a and 7b. There we see that BARTLETT with first order prewhitening has thin tails. For example, in Table 7a, the upper 5% tail area is equal to 2.2 for BARTLETT and equal to 7.8 for NW when $T = 1000$. When $T = 5000$, these numbers for BARTLETT and NW are respectively 2.6 and 5.2. The presence of thin tails reflects that the bandwidth parameter increases quite slowly as the number of observations increases. For example, for $T = 1000$ observations, Table 7a indicates that the bandwidth parameter is only 10.7 for BARTLETT. Figure 6 indicates that a bandwidth parameter of 10.7 could still produce a 30% overestimate of $\widehat{V}_{F,T}$. At the same time, the downward bias in estimates of $\widehat{V}_{F,T}$ disappears in a sample of 5000. Then the only source of bias that remains is the upward bias stemming from the overly short bandwidth parameter. This example suggests that the upward and downward sources of bias in this model are not perfectly correlated.

7 Summary and Conclusion

It is common in business cycle analysis to characterize the business cycle in terms of a set of second moment properties of detrended data. Researchers then focus on constructing models that can account for these moments. Recently, a number of researchers have employed versions of the GMM sampling theory developed by Hansen (1982) to assist them in this process. Although the sampling theory is known to be valid in arbitrarily large data sets, little is known about how well it works in realistic settings. Our purpose in this paper is to investigate this.

The results are disappointing. The asymptotic theory appears to provide a poor approximation in finite samples, particularly when the data have been HP filtered. These conclusions are based on a Monte Carlo study of two types of statistics. First, we examine second moment statistics like correlations and standard deviations. We study the coverage probabilities of confidence intervals computed for these statistics. Second, we study the size properties of a statistic proposed in Christiano and Eichenbaum (1992) for the purpose of testing model fit.

Consider first our analysis of second moment statistics. We study these in the context of two dgm's: a univariate time series model often used in the analysis of macroeconomic data and a four variable vector autoregression estimated for the basic macroeconomic variables using postwar U.S. data. Our objective in selecting these dgm's was to build confidence that our Monte Carlo results provide a reasonable indication of statistical performance in actual research applications. In our analysis of confidence intervals we consider two issues: (i) how frequently the 'true' value of the second moment lies outside computed confidence intervals and (ii) how often the true statistic lies to the left or to the right of the confidence interval. When the statistic lies outside the confidence interval more often than predicted by the (asymptotic) sampling theory, we say it exhibits a 'fat-tail' problem. Also, if it lies

on one side of the confidence interval more often than on the other side, then we say the statistic exhibits a 'skewness' problem. We examine a number of different ways of estimating confidence intervals and find that, for the 23 statistics examined, there is almost always a substantial skewness problem. Unfortunately, there is no clear pattern in the direction of this skewness problem. To a somewhat lesser extent, we also encountered a fat-tail problem.

We use our univariate dgm to conduct a thorough diagnosis of the causes of the skewness and fat tails. We find that these problems are most severe when the underlying detrended data exhibit substantial persistence. This is because our procedures for computing confidence intervals require an estimate of the zero-frequency spectral density of the 'GMM error process,' which is a particular function of the detrended data. As is well known, zero-frequency spectral density estimators show considerable imprecision when the data are persistent.

That persistence is an important element of the problem is consistent with our finding that the distortions are greater when we consider statistics based on data transformed using the HP filter, rather than the first difference filter. In our statistical environment, HP filtered data exhibit more persistence than first difference filtered data for two reasons: (i) the data have the property that there is more persistence in the business cycle frequencies than there is in the higher frequencies, and (ii) the HP filter emphasizes the former and the first difference filter, the latter. Among the various procedures we used, we found none that satisfactorily resolves the persistence problem. For example, there is a zero-frequency spectral density estimator designed specifically to accommodate persistence: the prewhitening procedure of Andrews (1991) and Andrews and Monahan (1992). However, we had only limited success with it. We did find that prewhitening can reduce (without eliminating) the fat-tail problem, but that this depends very sensitively on precisely how the degree of prewhitening is selected. In particular, we find that with first order prewhitening, the fat-tail problem is definitely alleviated relative to the situation when there is no prewhitening. However, this is only of limited interest: in any given empirical application, one does not know what the correct degree of prewhitening is, and some means must be found to estimate it. In our Monte Carlo analysis, we found that when the degree is selected using the AIC criterion, then second order prewhitening was most often indicated. The trouble is that second order prewhitening generates roughly the same results as no prewhitening at all. In addition, although there is some evidence that prewhitening can alleviate the fat-tail problem, it seems to have relatively little impact on skewness.

An important question addressed by our work is, What procedure for computing confidence intervals works best? Although we tried several procedures, none turned out to uniformly dominate the rest. The ones we tried are differentiated according to how the zero-frequency spectral density of the GMM error process is estimated. In each case, we consider the non-parametric approach, which involves computing a weighted average of autocovariances up to some finite lag length. In terms of sampling performance, the differences among the various procedures studied came down to how the lag length was chosen. One lag selection procedure, due to Andrews (1991) and Andrews and Monahan (1992), places structure on the autocovariances and computes the lag length based on an examination of the lag-one autocovariance. The other lag selection procedure, due to Newey and West (1994),

examines a longer list of autocovariances. Not surprisingly, we found that the Newey and West (1994) lag-selection procedure works best when the autocovariance function exhibits a complicated pattern, and the first order autoregressive assumption underlying the Andrews (1991) and Andrews and Monahan (1992) procedure is misspecified. We expected, therefore, that confidence intervals computed based on the Newey and West procedure would exhibit fewer distortions. We particularly expected this when there is first order prewhitening, since we found in our application that this produces an exotically shaped autocovariance function that is completely unlike the autocovariance function of a first order autoregression. It was to our initial great surprise, therefore, that with first order prewhitening the Andrews and Andrews and Monahan procedure actually dominates the Newey and West procedure. The reason for this is that small sample biases in the estimation of autocovariances also matter, in addition to misspecification, in determining finite sample performance. In our environment, the impact of these two factors roughly cancel. Since Andrews and Andrews and Monahan suffers from both problems and Newey and West only from only one, this explains why the former does better. Given the reason that the former dominated the latter in this case, we thought this experiment did not constitute a clear basis for preferring the Andrews and Andrews and Monahan procedure. This and other experiments left us with conflicting evidence on whether one or the other of these two procedures dominates.

Our basic findings for the univariate dgm extend to the multivariate dgm. In particular, for virtually all of the statistics examined, there is a skewness problem. Also, researchers who somehow know that first order prewhitening is appropriate, can alleviate the fat-tail problems somewhat. Moreover, the direction of skewness is not consistent across the 23 statistics examined.

Finally, we considered a statistic for testing the null hypothesis that a model's second moment implications correspond to the actual second moment properties of the data. We found that this statistic, which according to asymptotic sampling theory has a chi-square distribution, rejects the null hypothesis far too often, even when it is true. Again, the problem is more severe when the second moments are based on HP filtered data.

The results reported in this paper are a clear indication that a more reliable sampling theory is required for the statistics used in the analysis of business cycles. The need is less pressing for analysts who use the first difference filter. However, this is little comfort for researchers interested in the frequencies emphasized by the HP filter.

8 References

1. Andrews, Donald W. K., 1991, 'Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,' *Econometrica* 59 (3): 817–858, May.

2. Andrews, Donald W. K., and J. Christopher Monahan, 1992, 'An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator,' *Econometrica* 60: 953–966.

3. Backus, Dave, Allan Gregory, and Stanley Zin, 1989, 'Risk Premiums in the Term Structure: Evidence from Artificial Economies,' *Journal of Monetary Economics* 24.

4. Backus, Dave, and Patrick Kehoe, 1992, 'International Evidence on Historical Properties of Business Cycles,' *American Economic Review* 81, September.

5. Backus, Dave K., Patrick J. Kehoe, and Finn E. Kydland, 1994, 'Dynamics of the Trade Balance and the Terms of Trade: The J-Curve?' *American Economic Review* 84 (1): 84–103.

6. Baxter, Marianne, and Robert G. King, 1994, 'Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series,' National Bureau of Economic Research Working Paper 5022.

7. Braun, R. Anton, 1994, 'Tax Disturbances and Real Economic Activity in the Postwar United States,' *Journal of Monetary Economics* 33: 441–462.

8. Braun, R. Anton, and Charles L. Evans, 1991, 'Seasonal Solow Residuals and Christmas: A Case for Labor Hoarding and Increasing Returns,' August, manuscript.

9. Braun, R. Anton, and Charles L. Evans, 1995, 'Seasonality and Equilibrium Business Cycle Theories,' *Journal of Economic Dynamics and Control* 19 (3): 503–531.

10. Burnside, Craig, 1991, 'Small Sample Properties of 2-Step Method of Moments Estimators,' Department of Economics, University of Pittsburgh.

11. Burnside, Craig, and Martin Eichenbaum, 1994, 'Small Sample Properties of Generalized Method of Moments Based Wald Tests,' Working Paper, University of Pittsburgh and Northwestern University.

12. Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo, 1993, 'Labor Hoarding and the Business Cycle,' *Journal of Political Economy* 101 (2): 245–273.

13. Cecchetti, Stephen G., Pok-sang Lam, and Nelson C. Mark, 1993, 'The Equity Premium and the Risk-Free Rate: Matching the Moments,' *Journal of Monetary Economics* 31 (1): 21–46.

14. Christiano, Lawrence J., 1988, 'Why Does Inventory Investment Fluctuate So Much?,' *Journal of Monetary Economics* 21: 247–280.

15. Christiano, Lawrence J., 1989, 'Comment on Campbell and Mankiw,' *Macroeconomics Annual.*

16. Christiano, Lawrence J., 1992, 'Searching for a Break in GNP,' *Journal of Business and Economic Statistics* 10: 3–23.

17. Christiano, Lawrence J., and Martin Eichenbaum, 1992, 'Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations,' *American Economic Review* 82 (3): 430–450.

18. Cogley, Timothy, and James M. Nason, 1995, 'Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research,' *Journal of Economic Dynamics and Control* 19 (1&2): 253–278.

19. Cooley, Thomas F., and Gary D. Hansen, 1989, 'The Inflation Tax in a Real Business Cycle Model,' *American Economic Review* 79 (4): 733–748.

20. Davidson, Russel, and James G. MacKinnon, 1993, *Estimation and Inference in Econometrics*, Oxford University Press.

21. Den Haan, Wouter J., 1995, 'The Term Structure of Interest Rates in Real and Monetary Economies,' *Journal of Economic Dynamics and Control.*

22. Den Haan, Wouter J., and Andrew Levin, 1994, 'Inferences from Parametric and Non-Parametric Spectral Density Estimation Procedures,' manuscript, UCSD and Federal Reserve Board.

23. Den Haan, Wouter J., and Albert Marcet, 1994, 'Accuracy in Simulations,' *Review of Economic Studies* 61 (1): 3–18.

24. Ericsson, Neil R., 1991, 'Monte Carlo Methodology and the Finite Sample Properties of Instrumental Variables Statistics for Testing Nested and Non-Nested Hypotheses, *Econometrica* 59 (5): 1249–1277.

25. Ferson, Wayne, and Stephen E. Foerster, 1991, 'Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models,' Working Paper, University of Chicago Graduate School of Business.

26. Fisher, Jonas D. M., 1993, 'Relative Prices, Complementarities, and Co-movement Among Components of Aggregate Expenditures,' Working Paper, University of Western Ontario.

27. 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, Washington, D. C., April.

28. Hamilton, James D., 1994, *Time Series Analysis*, Princeton University Press.

29. Hansen, Lars P., 1982, 'Large Sample Properties of Generalized Method of Moment Models,' *Econometrica* 50: 1269–1286.

30. Hansen, Lars P., and Robert J. Hodrick, 1980, 'Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis,' *Journal of Political Economy* 88: 829–853.

31. King, Robert G., and Sergio T. Rebelo, 1993, 'Low Frequency Filtering and Real Business Cycles,' *Journal of Economic Dynamics and Control* 17 (1/2): 207–232.

32. Kocherlakota, N. R., 1990, 'On Tests of Representative Consumer Asset Pricing Models, *Journal of Monetary Economics* 26 (2): 285–304.

33. Kydland, Finn E., and Edward C. Prescott, 1982, 'Time to Build and Aggregate Fluctuations,' *Econometrica* 50 (6): 1345–1370.

34. Long, John B., and Charles I. Plosser, 1982, 'Real Business Cycles,' *Journal of Political Economy* 91 (1): 39–69.

35. Marshall, David, 1992, 'Inflation and Asset Returns in a Monetary Economy,' *Journal of Finance* 47 (4): 1315–1342.

36. Neely, Christopher J., 1993, 'A Reconsideration of Representative Consumer Asset Pricing Models,' Department of Economics, University of Iowa.

37. Nelson, Charles R., and Richard Startz, 1990, 'The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One,' *Journal of Business* 63 (1, Part 2): S125–S140.

38. Newey, W. K., and K. D. West, 1987, 'A Simple Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,' *Econometrica* 55: 703–708.

39. Newey, W. K., and K. D. West, 1994, 'Automatic Lag Selection in Covariance Matrix Estimation,' *Review of Economic Studies* 61 (4): 631–653.

40. Ogaki, Masao, 1992, 'An Introduction to the Generalized Method of Moments,' Rochester Center for Economic Research no. 370.

41. Prescott, Edward, 1986, 'Theory Ahead of Business Cycle Measurement,' *Carnegie-Rochester Conference Series on Public Policy* 25: 11–66.

42. Reynolds, Patricia, 1992, 'International Co-Movements in Production and Consumption' Working Paper, University of Southern California.

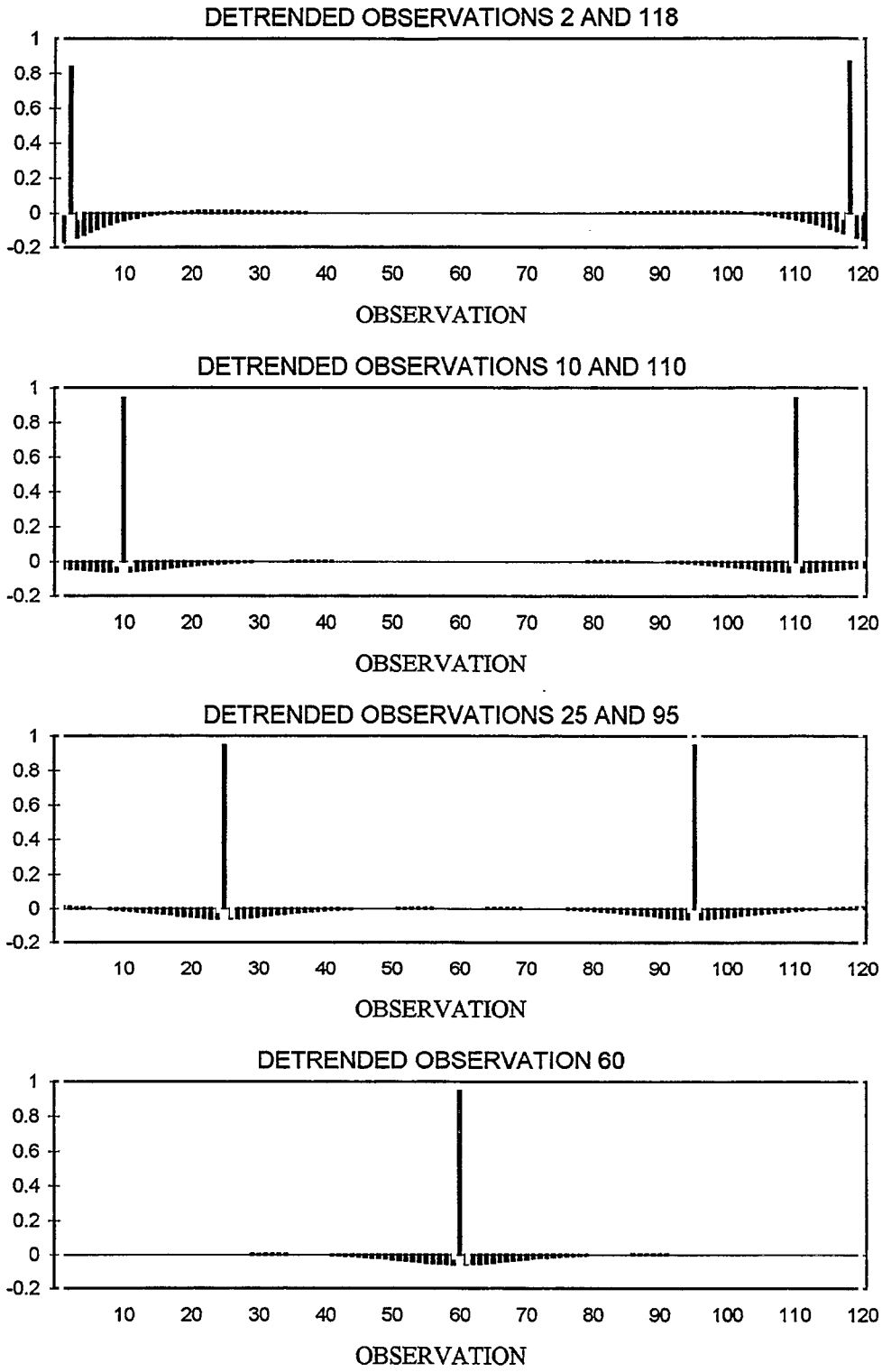
43. Singleton, K. J., 1988, 'Econometric Issues in the Analysis of Equilibrium Business Cycle Models,' *Journal of Monetary Economics* 21 (2/3): 361–386.

44. Tauchen, G., 1986, 'Statistical Properties of Generalized Method of Moment Models of Structural Parameters Obtained from Financial Market Data,' *Journal of Business and Economic Statistics* 4: 397–425.

45. West, Ken, and David W. Wilcox, 1992, 'Some Evidence on Finite Sample Distributions of Instrumental Variables Estimators of a Linear-Quadratic Inventory Model,' Board of Governors, Federal Reserve System.

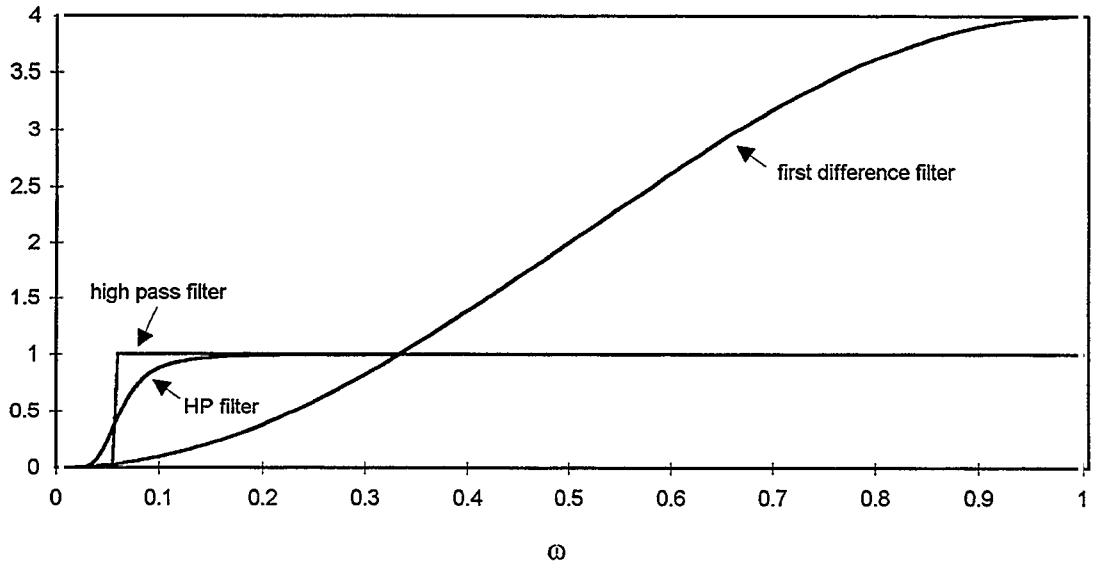
46. White, H., 1984, *Asymptotic Theory for Econometricians*, New York: Academic Press.

FIGURE 1a: FILTER WEIGHTS, HP FILTER



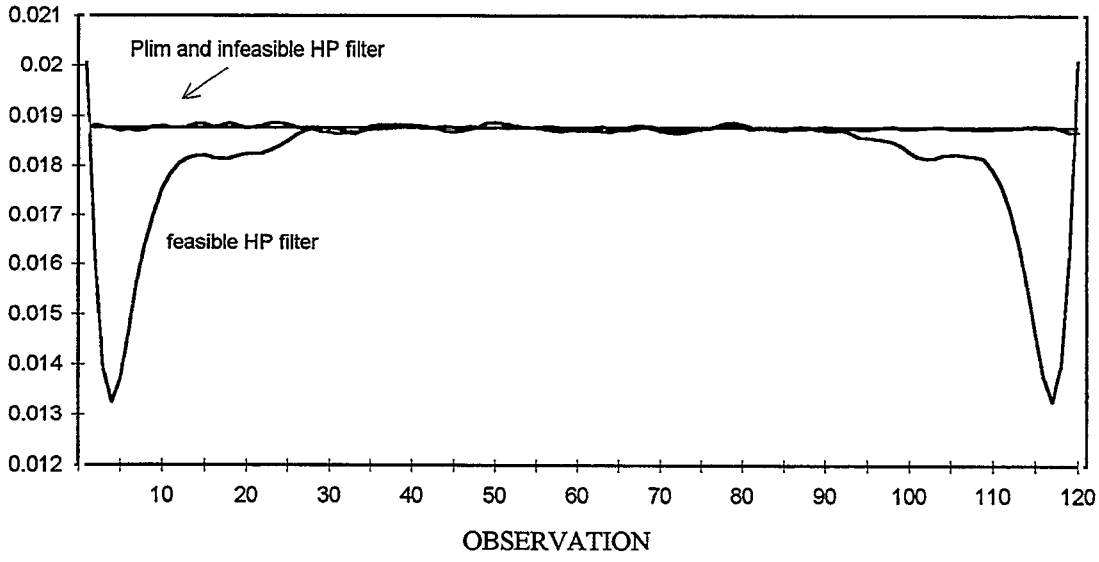
NOTES: The figure displays the HP filter weights used to compute the i-th observation on the detrended variable, as indicated in the header. The sample length is equal to 120 observations.

FIGURE 1b: TRANSFER FUNCTIONS, ALTERNATIVE FILTERS



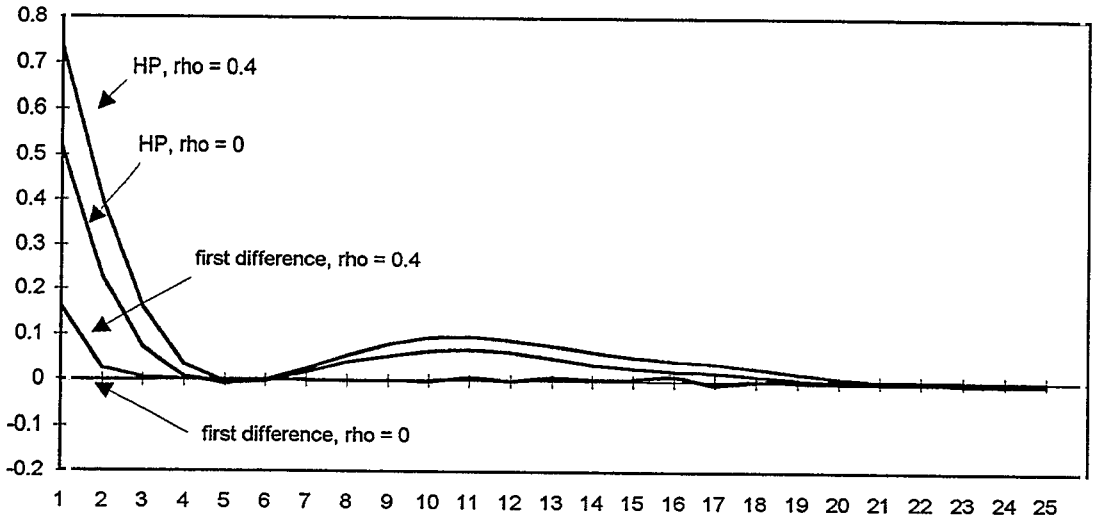
Note: ω denotes frequency of oscillations divided by π.

FIGURE 1c: STANDARD DEVIATION OF HP FILTERED DATA



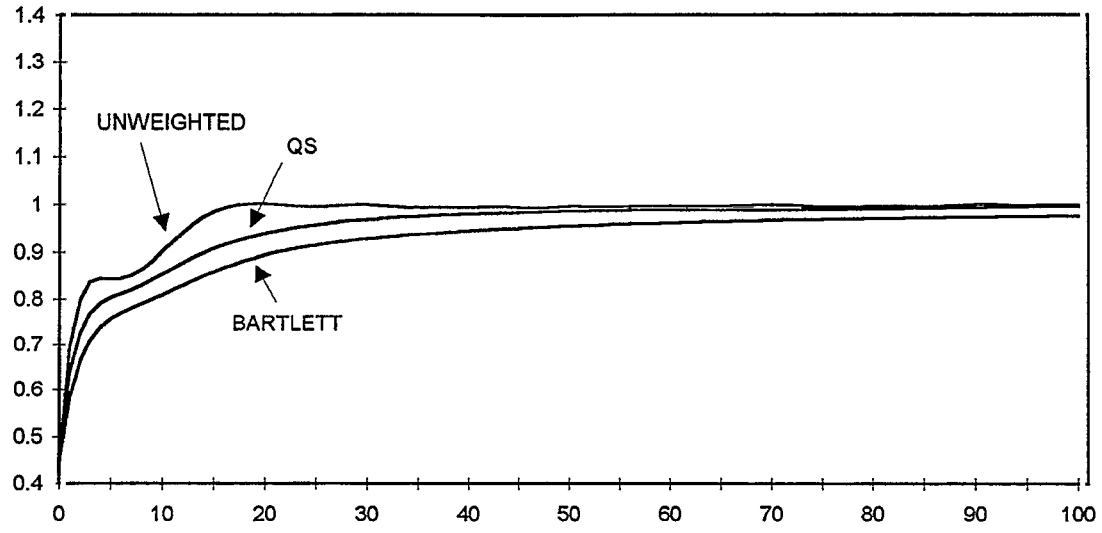
Notes: Plot of the standard deviation of x_t^d, where x_t^d is obtained by detrending 120 observations using the indicated filter. Plim signifies $\underset{T\to\infty}{P\lim}\hat{\psi}_T$. The dgm is equation (30) with $\rho = 0.4$ and $\sigma = 0.01$. See section 5.1.1 for details.

FIGURE 1d: AUTOCORRELATION COEFFICIENTS OF $\left(x_t^d\right)^2$



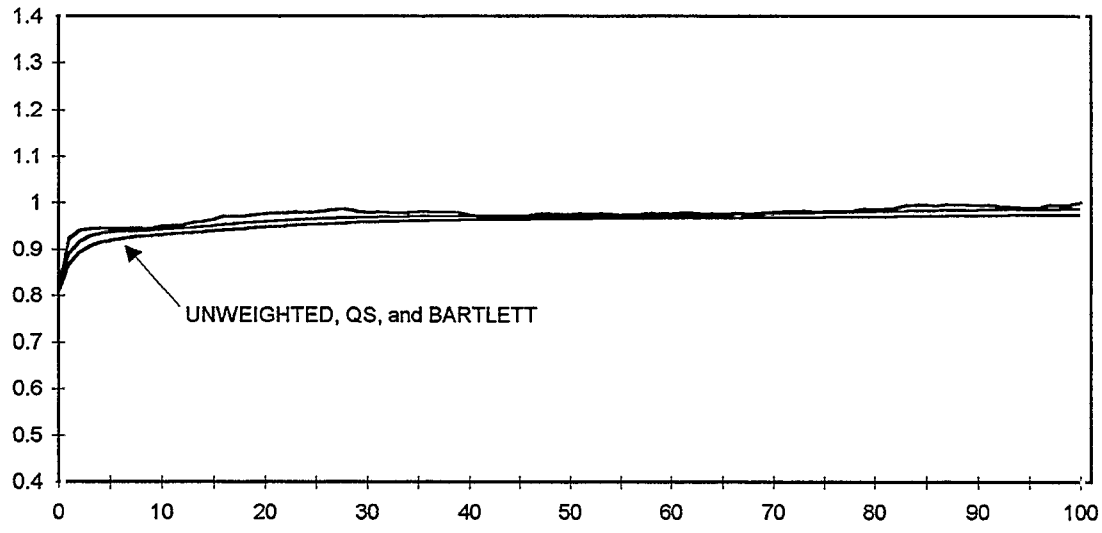
Notes: The figure plots the autocorrelation coefficients of $\left(x_t^d\right)^2$ when the raw data are generated by equation (30) with $\rho = 0.4$ and $\sigma = 0.01$.

FIGURE 2a: $S(\xi)$ divided by S, HP filter and $b = 0$.



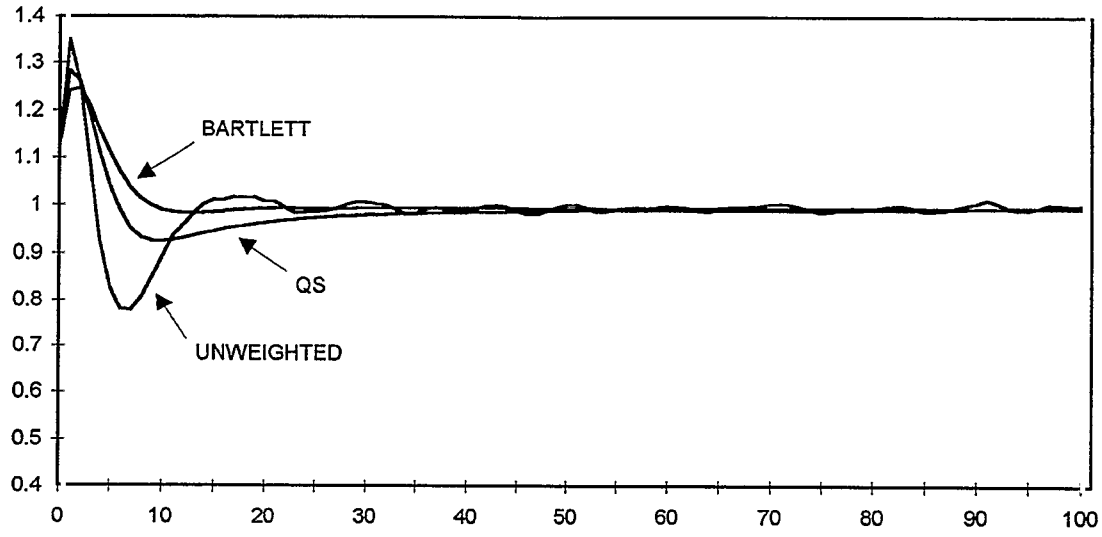
Notes: $S(\xi)$ is the spectral density of the GMM residual associated with data detrended by the indicated filter, using the indicated kernel and bandwidth ξ. S is equal to $S(\xi)$ as $\xi \to \infty$. The dgm is equation (30) with $\rho = 0.4$ and $\sigma = 0.01$. See Table 1 for definitions of the spectral estimators.

FIGURE 2b: $S(\xi)$ divided by S, first difference filter and $b = 0$.



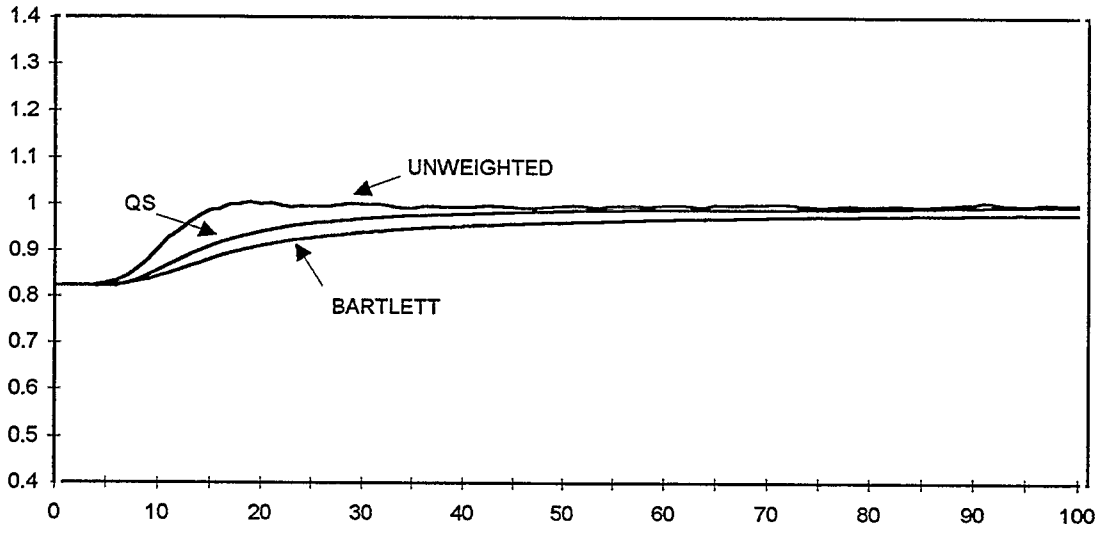
Note: See the note to Figure 2a.

FIGURE 3: $S(\xi)$ divided by S, HP filter and $b = 1$.



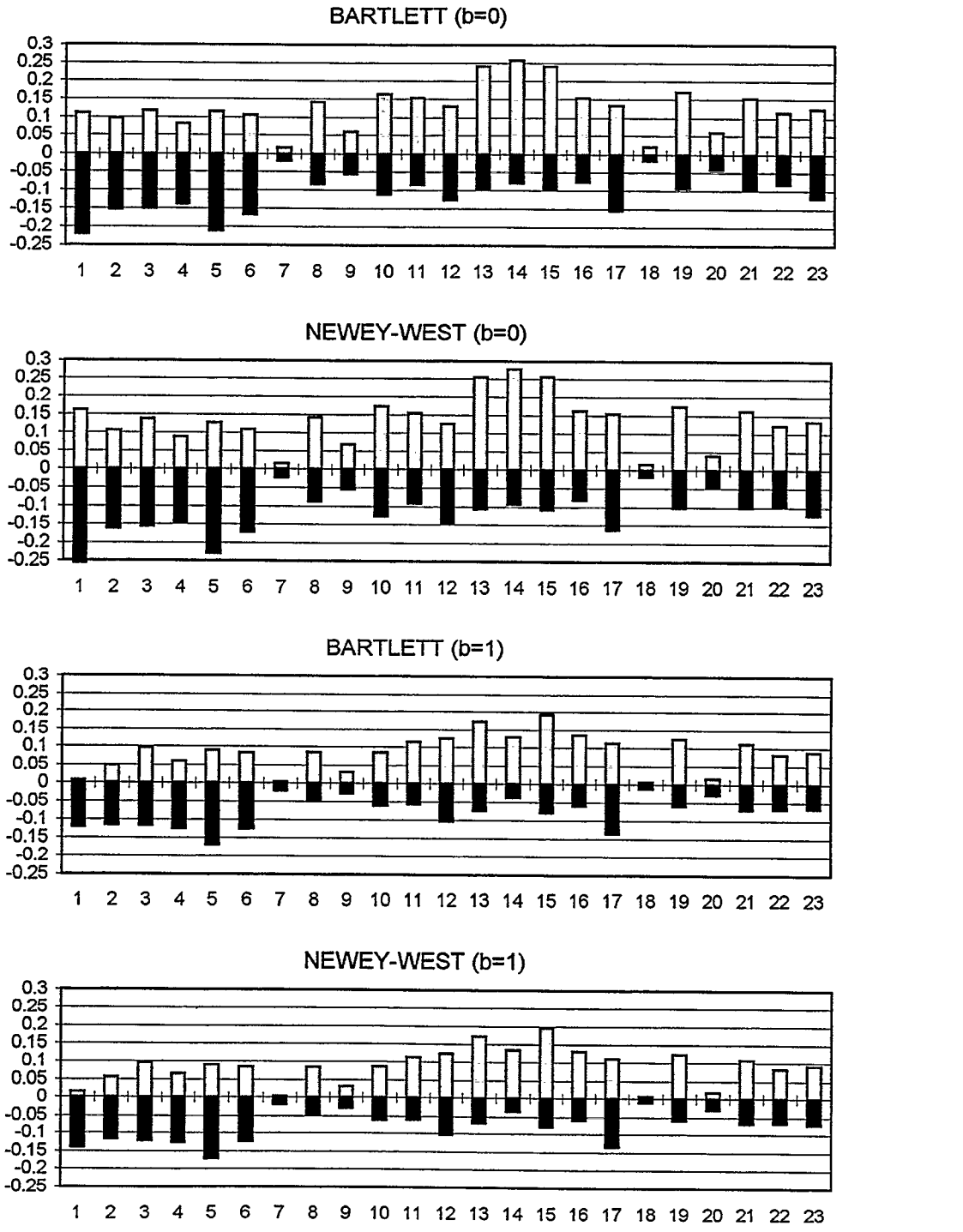
Note: See the note to Figure 2a.

FIGURE 4: $S(\xi)$ divided by S, HP filter and $b = 2$.



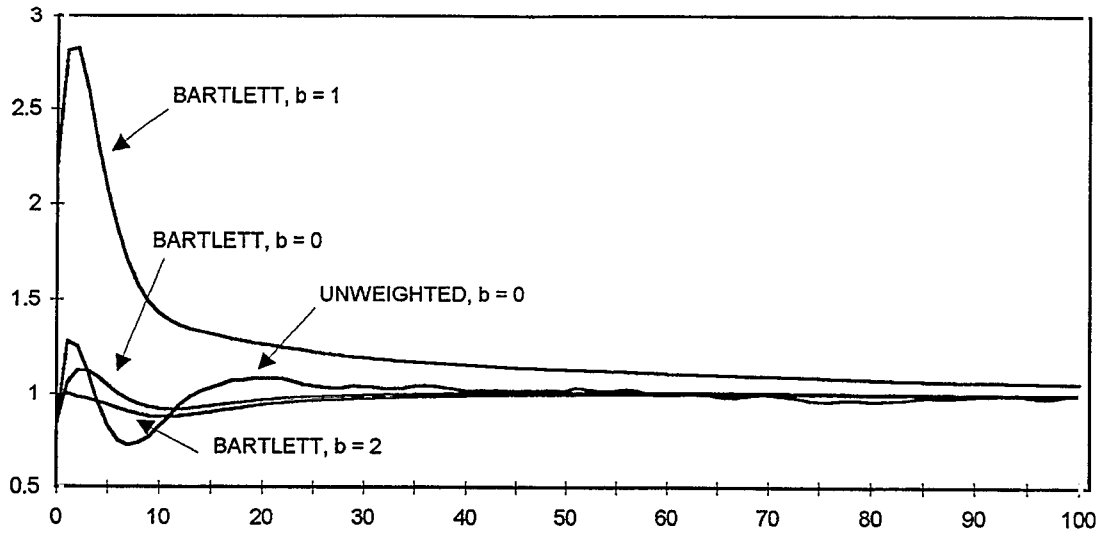
Note: See the note to Figure 2a.

FIGURE 5: COVERAGE PROBABILITIES FOR VARIOUS STATISTICS (MULTIVARIATE DGM)



Notes: The grey(black) columns report the frequency the t-statistic is higher(less) than the upper(lower) 5% critical value. The value of b indicates the order of prewhitening. For definitions of the spectral estimators see Table 1. The ordering of the statistics is as follows: 1 = σ_y , 2 = σ_c/σ_y , 3 = σ_i/σ_y , 4 = σ_n/σ_y, 5 = σ_w/σ_y , 6 = σ_w/σ_n , 7 = $\rho_{yy}(-1)$, 8 = $\rho_{cy}(-1)$, 9 = $\rho_{iy}(-1)$, 10 = $\rho_{ny}(-1)$, 11 = $\rho_{wy}(-1)$, 12 = $\rho_{wn}(-1)$, 13 = $\rho_{cy}(0)$, 14 = $\rho_{iy}(0)$, 15= $\rho_{ny}(0)$, 16 = $\rho_{wy}(0)$, 17 = $\rho_{wn}(0)$, 18 = $\rho_{yy}(1)$, 19 = $\rho_{cy}(1)$, 20 = $\rho_{iy}(1)$, 15= $\rho_{ny}(1)$, 16 = $\rho_{wy}(1)$, 17 = $\rho_{wn}(1)$, where σ_x stands for the standard deviation of variable x, and ρ_{zx} (τ) stands for the correlation coefficient between z_t and $x_{t+\tau}$.

FIGURE 6: $S(\xi)$ divided by S, Wald Test, experiment 2 and the HP filter.



Notes: $S(\xi)$ is the spectral density of the GMM residuals in experiment 2, using the indicated kernel and bandwidth ξ. S is equal to $S(\xi)$ as $T \to \infty$. For details see section 6.2.2, and see Table 1 for definitions of the spectral estimators.

TABLE 1: ZERO-FREQUENCY SPECTRAL DENSITY ESTIMATORS

NAME	KERNEL		BANDWIDTH
UW(11)	UNWEIGHTED	(EQUATION (18) WITH $\theta = 0$)	11
BART(11)	BARTLETT	(EQUATION (18) WITH $\theta = 1$)	11
BARTLETT	BARTLETT	(EQUATION (18) WITH $\theta = 1$)	ANDREWS, AUTOMATIC
QS	QS	(EQUATION (19) AND (20))	ANDREWS, AUTOMATIC
NW	BARTLETT	(EQUATION (18) WITH $\theta = 1$)	NEWEY-WEST, AUTOMATIC

TABLE 2a: COVERAGE PROBABILITIES

DGM: $\Delta x_t = 0.4\,\Delta x_{t-1} + 0.01\varepsilon_t, \quad \varepsilon_t \sim N(0,1)$

SPECTRAL ESTIMATOR	b	HP 5%	HP 10%	HP 90%	HP 95%	Δ 5%	Δ 10%	Δ 90%	Δ 95%
					T = 120				
TRUE		3.6	9.2	10.6	6.6	4.7	9.5	10.0	4.8
UW(11)	0	18.7	23.8	15.2	10.8	12.6	17.7	13.0	8.1
BART(11)	0	19.3	23.6	16.7	10.3	11.0	16.1	10.8	5.5
BARTLETT	0	19.0	23.2	16.4	9.8	9.5	14.7	9.9	4.6
QS	0	18.6	22.9	15.0	9.4	9.5	14.4	9.3	4.2
NW	0	20.9	24.4	18.0	11.7	10.0	15.1	10.7	5.4
UW(11)	1	19.6	23.9	16.0	11.0	13.1	17.8	13.4	8.1
BART(11)	1	16.5	21.1	9.9	5.8	10.9	16.0	10.3	5.1
BARTLETT	1	12.1	17.5	5.5	1.6	8.5	13.7	8.1	3.2
QS	1	12.0	17.3	5.0	1.3	8.8	14.0	8.2	3.4
NW	1	16.7	20.5	9.3	5.1	9.0	14.1	9.5	4.4
UW(11)	2	19.1	24.0	15.8	10.8	13.3	17.7	13.4	8.3
BART(11)	2	18.6	22.8	15.8	9.8	11.1	15.8	10.5	5.5
BARTLETT	2	18.8	22.3	15.5	9.6	9.5	14.1	8.8	4.0
QS	2	18.8	22.2	15.7	9.7	9.7	14.2	8.9	4.1
NW	2	18.8	22.1	15.5	9.6	9.7	14.5	9.1	5.1
					T = 1000				
TRUE		4.5	9.1	10.7	5.5	4.9	10.0	9.7	5.3
UW(11)	0	8.8	15.7	13.3	7.7	7.8	13.3	10.6	5.2
BART(11)	0	11.1	17.4	15.4	9.4	7.8	13.3	10.3	5.5
BARTLETT	0	9.0	15.5	13.0	7.1	7.8	13.6	10.3	5.9
QS	0	8.6	15.4	12.6	6.8	7.7	13.4	10.0	5.6
NW	0	11.1	17.7	15.2	9.0	7.9	13.5	10.3	6.2
UW(11)	1	9.6	16.2	13.2	8.4	7.9	13.4	10.6	5.3
BART(11)	1	7.4	13.0	10.0	4.9	7.7	13.1	9.8	5.3
BARTLETT	1	5.7	10.7	7.7	3.9	7.2	13.1	9.5	5.2
QS	1	5.0	9.3	6.8	3.1	7.1	13.1	9.5	5.2
NW	1	7.5	13.6	11.1	5.3	7.0	12.9	9.6	5.3
UW(11)	2	9.0	15.8	13.1	7.5	7.9	13.4	10.6	5.3
BART(11)	2	10.0	17.2	14.8	8.6	7.5	13.1	10.1	5.2
BARTLETT	2	10.8	17.4	15.0	9.2	7.4	13.1	9.5	5.4
QS	2	10.8	17.3	15.1	9.1	7.4	13.1	9.5	5.4
NW	2	10.7	17.3	15.0	9.1	7.0	12.9	9.6	5.3

NOTES: The 5%(95%) and 10%(90%) columns report the frequency the t-statistic is less(higher) than the lower(upper) 5% and 10% critical value. The row true uses the Monte Carlo standard deviation of the parameter estimate to calculate the standard error. b indicates the order of prewhitening. For definitions of the spectral estimators see Table 1. HP(Δ) refers to Hodrick-Prescott detrending (first-differencing). Based on 1,000 independently simulated data sets, each of length T.

TABLE 2b: COVERAGE PROBABILITIES

DGM: $\Delta x_t = 0.1\Delta x_{t-1} + 0.01\varepsilon_t, \quad \varepsilon_t \sim N(0,1)$

SPECTRAL ESTIMATOR	b	HP 5%	HP 10%	HP 90%	HP 95%	Δ 5%	Δ 10%	Δ 90%	Δ 95%
					T = 120				
TRUE		3.7	9.0	10.5	6.5	4.9	9.8	10.4	5.2
UW(11)	0	18.9	24.2	14.7	10.6	11.6	15.6	12.8	8.8
BART(11)	0	18.0	24.0	15.4	9.4	9.7	14.2	10.4	5.9
BARTLETT	0	17.9	23.9	15.4	9.1	7.9	13.0	9.7	4.0
QS	0	17.8	23.1	15.0	8.9	7.8	13.2	9.8	3.8
NW	0	19.1	24.7	17.1	10.8	9.2	13.6	10.2	4.8
UW(11)	1	19.6	23.9	15.3	11.5	11.7	15.8	12.7	8.5
BART(11)	1	16.7	21.4	12.3	7.3	9.6	13.9	10.7	5.6
BARTLETT	1	14.8	19.1	8.5	4.0	7.7	12.5	9.6	4.0
QS	1	14.7	19.0	8.1	4.1	7.7	13.0	9.7	4.0
NW	1	16.2	20.6	11.0	6.0	9.2	13.4	10.1	4.9
UW(11)	2	19.1	23.8	15.2	11.5	11.8	15.6	12.9	8.7
BART(11)	2	17.5	22.9	14.1	8.8	9.8	14.3	10.5	5.9
BARTLETT	2	16.6	21.6	13.1	7.7	8.4	12.9	9.2	4.2
QS	2	16.6	21.9	13.3	7.8	8.6	13.1	9.4	4.3
NW	2	16.9	22.0	13.4	7.7	9.0	13.7	9.8	5.4
					T = 1000				
TRUE		4.3	8.7	10.6	5.8	4.7	9.5	10.2	5.4
UW(11)	0	8.3	16.1	13.0	6.7	6.9	13.0	11.5	5.3
BART(11)	0	10.5	17.8	15.0	8.4	6.8	12.4	10.9	5.7
BARTLETT	0	9.3	17.0	13.5	7.1	7.0	12.2	10.5	5.4
QS	0	9.5	17.2	13.5	7.0	6.8	12.2	10.5	5.4
NW	0	11.1	18.1	14.7	8.6	7.0	12.4	10.8	5.6
UW(11)	1	8.3	15.9	13.1	7.0	6.9	13.0	11.5	5.3
BART(11)	1	8.5	15.2	12.4	6.2	6.8	12.4	10.8	5.6
BARTLETT	1	7.1	12.9	9.7	4.5	6.7	12.1	10.4	5.3
QS	1	4.7	9.7	11.2	6.3	6.8	12.1	10.4	5.3
NW	1	8.3	14.8	12.1	6.2	6.9	12.4	10.7	5.5
UW(11)	2	8.3	16.0	13.0	6.8	7.0	13.0	11.5	5.3
BART(11)	2	8.9	16.6	13.3	7.1	6.7	12.5	10.9	5.6
BARTLETT	2	9.3	16.1	13.2	7.0	7.1	12.4	10.6	5.2
QS	2	9.3	16.1	13.2	7.0	7.1	12.4	10.7	5.2
NW	2	9.3	16.5	13.6	7.2	7.1	12.5	10.6	5.5

For notes see Table 2a.

TABLE 3a: DIAGNOSING FAT-TAIL AND SKEWNESS PROBLEM IN TABLE 2a

DGM: $\Delta x_t = 0.4\,\Delta x_{t-1} + 0.01\varepsilon_t, \quad \varepsilon_t \sim N(0,1)$

		HP			Δ		
SPECTRAL ESTIMATOR	b	I	II	III	I	II	III
				T = 120			
UW(11)	0	.50	.00269	.001870	.34	.000827	.000680
BART(11)	0	.62	=	.001790	.47	=	.000725
BARTLETT	0	.61	=	.001810	.54	=	.000761
QS	0	.59	=	.001867	.53	=	.000772
NW	0	.65	=	.001687	.49	=	.000740
UW(11)	1	.33	=	.001973	.32	=	.000680
BART(11)	1	.48	=	.002277	.47	=	.000738
BARTLETT	1	.56	=	.002829	.52	=	.000808
QS	1	.57	=	.002867	.52	=	.000805
NW	1	.47	=	.002346	.47	=	.000782
UW(11)	2	.48	=	.001867	.33	=	.000678
BART(11)	2	.59	=	.001862	.45	=	.000735
BARLETT	2	.62	=	.001871	.46	=	.000793
QS	2	.62	=	.001864	.46	=	.000789
NW	2	.62	=	.001869	.45	=	.000781
				T = 1000			
UW(11)	0	.46	.000959	.000819	.37	.000299	.000280
BART(11)	0	.55	=	.000741	.47	=	.000278
BARTLETT	0	.49	=	.000826	.53	=	.000276
QS	0	.50	=	.000835	.53	=	.000279
NW	0	.48	=	.000751	.41	=	.000274
UW(11)	1	.36	=	.000809	.36	=	.000280
BART(11)	1	.59	=	.000911	.47	=	.000282
BARTLETT	1	.58	=	.000987	.55	=	.000285
QS	1	.58	=	.001045	.55	=	.000285
NW	1	.45	=	.000898	.48	=	.000284
UW(11)	2	.46	=	.000820	.36	=	.000280
BART(11)	2	.53	=	.000771	.46	=	.000282
BARTLETT	2	.54	=	.000755	.51	=	.000285
QS	2	.54	=	.000755	.51	=	.000284
NW	2	.55	=	.000760	.47	=	.000283

For notes see Table 2a and
I: the correlation between $\hat{\psi}_T$ and $[\hat{V}_T]^{1/2}$.
II: the Monte-Carlo sd of $\hat{\psi}_T$.
III: the Monte-Carlo mean of $[\hat{V}_T]^{1/2}$.

TABLE 3b: DIAGNOSING FAT-TAIL AND SKEWNESS PROBLEMS IN TABLE 2b

DGM: $\Delta x_t = 0.1\,\Delta x_{t-1} + 0.01\varepsilon_t, \quad \varepsilon_t \sim N(0,1)$

SPECTRAL ESTIMATOR	b	HP I	HP II	HP III	Δ I	Δ II	Δ III
				T = 120			
UW(11)	0	.49	.00175	.001249	.24	.000663	.000559
BART(11)	0	.61	=	.001217	.39	=	.000601
BARTLETT	0	.65	=	.001210	.56	=	.000642
QS	0	.63	=	.001243	.55	=	.000641
NW	0	.61	=	.001140	.44	=	.000622
UW(11)	1	.43	=	.001251	.24	=	.000558
BART(11)	1	.62	=	.001358	.39	=	.000602
BARTLETT	1	.66	=	.001561	.50	=	.000644
QS	1	.66	=	.001574	.50	=	.000641
NW	1	.63	=	.001413	.41	=	.000623
UW(11)	2	.46	=	.001241	.24	=	.000555
BART(11)	2	.60	=	.001278	.37	=	.000600
BARLETT	2	.65	=	.001324	.42	=	.000636
QS	2	.65	=	.001318	.42	=	.000634
NW	2	.64	=	.001315	.38	=	.000622
				T = 1000			
UW(11)	0	.47	.000628	.000541	.31	.000238	.000223
BART(11)	0	.56	=	.000496	.43	=	.000224
BARTLETT	0	.53	=	.000523	.53	=	.000226
QS	0	.54	=	.000520	.52	=	.000226
NW	0	.47	=	.000494	.44	=	.000225
UW(11)	1	.45	=	.000539	.31	=	.000223
BART(11)	1	.59	=	.000546	.42	=	.000225
BARTLETT	1	.63	=	.000599	.49	=	.000227
QS	1	.63	=	.000607	.49	=	.000227
NW	1	.56	=	.000555	.43	=	.000225
UW(11)	2	.47	=	.000541	.31	=	.000223
BART(11)	2	.55	=	.000523	.41	=	.000224
BARTLETT	2	.58	=	.000527	.45	=	.000226
QS	2	.58	=	.000527	.45	=	.000226
NW	2	.56	=	.000521	.41	=	.000225

For notes see Table 2a and

I: the correlation between $\hat{\psi}_T$ and $[\hat{V}_T]^{1/2}$.

II: the Monte-Carlo sd of $\hat{\psi}_T$.

III: the Monte-Carlo mean of $[\hat{V}_T]^{1/2}$.

TABLE 4a: SAMPLING PROPERTIES OF BANDWIDTHS

DGM: $\Delta x_t = 0.4\,\Delta x_{t-1} + 0.01\varepsilon_t, \quad \varepsilon_t \sim N(0,1)$

		HP		Δ	
SPECTRAL ESTIMATOR	b	I	II	I	II
			T = 120		
BARTLETT	0	10.7	2.85	2.41	1.31
QS	0	10.0	3.21	2.29	1.03
NW	0	5.0	1.74	4.42	3.96
BARTLETT	1	3.26	.84	.46	.36
QS	1	2.96	.45	.72	.34
NW	1	13.05	10.11	5.00	3.85
BARLETT	2	.71	.44	.32	.26
QS	2	.95	.40	.58	.27
NW	2	3.18	1.84	4.45	3.47
			T = 1000		
BARTLETT	0	24.26	2.11	5.30	1.08
QS	0	17.30	1.76	3.76	.60
NW	0	11.79	2.64	6.63	3.41
BARTLETT	1	6.91	.83	.45	.25
QS	1	4.67	.48	.72	.26
NW	1	40.70	14.86	6.75	3.76
BARTLETT	2	.74	.43	.17	.13
QS	2	.98	.37	.39	.18
NW	2	5.86	3.09	6.68	3.66

For notes see Table 2a and
I: the Monte-Carlo mean of ξ_T.
II: the Monte-Carlo sd of ξ_T.

TABLE 4b: SAMPLING PROPERTIES OF BANDWIDTHS

DGM: $\Delta x_t = 0.1\ \Delta x_{t-1} + 0.01856\varepsilon_t, \quad \varepsilon_t \sim N(0,1)$

		HP		Δ	
SPECTRAL ESTIMATOR	b	I	II	I	II
			T = 120		
BARTLETT	0	7.14	2.03	1.47	.78
QS	0	6.37	1.98	1.49	.56
NW	0	4.44	2.03	5.10	3.94
BARTLETT	1	1.52	.85	.31	.23
QS	1	1.62	.67	.57	.25
NW	1	7.74	6.30	5.08	3.72
BARLETT	2	.41	.30	.30	.21
QS	2	.67	.30	.56	.23
NW	2	3.87	2.43	4.73	3.14
			T = 1000		
BARTLETT	0	16.26	1.62	1.51	.80
QS	0	10.93	1.21	1.52	.54
NW	0	10.31	3.91	6.68	3.84
BARTLETT	1	2.84	.81	.15	.12
QS	1	2.38	.48	.37	.17
NW	1	16.10	7.21	6.69	3.82
BARTLETT	2	.41	.24	.15	.11
QS	2	.68	.26	.37	.16
NW	2	7.50	4.22	6.71	3.78

For notes see Table 2a and
I: the Monte-Carlo mean of ξ_T.
II: the Monte-Carlo sd of ξ_T.

TABLE 5a: COVERAGE PROABILITIES FOR σ_y

DGM: US VAR

			HP		Δ	
SPECTRAL ESTIMATOR	b	DGM	5%	95%	5%	95%
BART(11)	0	N	21.6	11.6	15.4	7.8
BART(11)	0	B	20.6	9.8	11.8	6.2
BARTLETT	0	N	22.2	11.0	14.0	7.2
BARTLETT	0	B	21.0	9.4	11.6	5.4
NW	0	N	25.6	16.0	14.4	8.2
NW	0	B	23.4	14.2	12.0	6.2
BART(11)	1	N	10.6	0.6	15.2	7.8
BART(11)	1	B	9.8	0.2	11.6	4.8
BARTLETT	1	N	12.2	0.6	13.8	7.2
BARTLETT	1	B	11.2	0.4	11.4	4.2
NW	1	N	14.0	1.6	13.8	7.2
NW	1	B	11.2	1.0	11.4	4.2

NOTES: The 5%(95%) column reports the frequency the t-statistic is less(higher) than the lower(upper) 5% critical value. The value of b indicates the order of prewhitening. The DGM is a VAR described in the test. The generated errors are either normal (DGM = N) or bootstrapped (DGM = B). For definitions of the spectral estimators see Table 1. HP refers to Hodrick-Prescott detrending and Δ refers to first-differencing.

TABLE 5b: COVARAGE PROBABILITES FOR VARIOUS STATISTICS, SUMMARY RESULTS

DGM: US VAR

			HP				Δ			
SPECTRAL ESTIMATOR	b	DGM	I	II	III	IV	I	II	III	IV
BART(11)	0	N	18.1	23	7.5	7	7.3	20	3.5	9
BART(11)	0	B	15.9	23	6.2	6	7.2	20	3.4	11
BARTLETT	0	N	14.8	21	6.0	8	6.0	16	3.4	10
BARTLETT	0	B	13.6	20	5.1	8	5.7	17	2.8	14
NW	0	N	20.7	20	5.9	11	9.9	16	3.4	8
NW	0	B	19.3	20	5.3	9	9.5	17	2.9	13
BART(11)	1	N	10.1	23	5.6	7	6.2	21	3.2	9
BART(11)	1	B	9.0	22	4.5	5	5.8	20	3.1	12
BARTLETT	1	N	8.7	19	4.8	10	5.2	17	3.1	10
BARTLETT	1	B	7.7	19	3.8	11	4.9	17	2.8	12
NW	1	N	12.2	19	4.8	10	8.6	17	3.1	10
NW	1	B	11.2	19	3.7	11	7.9	17	2.8	12

Notes: See Table 5a, and

I: the average of the absolute deviation between the sum of the two 5% tail areas minus 10% across all 23 statistics.

II: The number of times the sum of the two 5% tail areas is larger than 10%.

III: the average absolute value of the difference between the upper and the lower 5% tail areas.

IV: the number of times the lower 5% tail area is bigger than the upper 5% tail area.

TABLE 6a: COVERAGE PROBABILITIES FOR WALD TEST EXAMPLE: EXPERIMENT 1

$\rho = 0.1$

HP FILTERED DATA

SPECTRAL ESTIMATOR	T	b	5%	10%	90%	95%	AVE(ξ_T)
TRUE	120	-	4.6	8.8	10.2	4.8	-
BARTLETT	120	0	3.5	6.9	28.5	20.5	10.4
NW	120	0	4.3	7.7	24.1	16.4	5.2
BARTLETT	120	1	5.0	10.6	17.7	11.9	3.2
NW	120	1	5.3	9.5	21.1	15.5	9.9
BARTLETT	120	2	3.8	7.0	26.3	19.8	.81
NW	120	2	4.2	8.0	26.0	17.9	3.1
TRUE	1000	-	5.3	9.0	10.1	5.6	-
BARTLETT	1000	0	4.7	8.3	16.6	11.0	23.3
NW	1000	0	3.6	6.9	16.2	10.0	12.8
BARTLETT	1000	1	5.5	9.6	11.3	7.0	7.0
NW	1000	1	3.8	8.3	12.3	7.6	29.7
BARTLETT	1000	2	4.8	8.0	17.6	11.7	.86
NW	1000	2	3.5	7.1	16.2	10.1	6.44
TRUE	5000	-	3.7	9.1	10.0	4.7	-
BARTLETT	5000	0	3.6	8.2	12.7	6.7	40.3
NW	5000	0	4.0	8.9	11.9	7.0	28.4
BARTLETT	5000	1	3.7	8.7	11.8	5.9	12.0
NW	5000	1	4.3	9.2	10.2	5.7	39.0
BARTLETT	5000	2	3.2	7.7	16.8	9.3	1.04
NW	5000	2	3.8	8.4	13.3	8.1	19.3

FIRST DIFFERENCED DATA

SPECTRAL ESTIMATOR	T	b	5%	10%	90%	95%	AVE(ξ_T)
TRUE	120	-	4.5	9.4	9.9	4.9	-
BARTLETT	120	0	3.7	9.1	11.2	6.1	2.3
BARTLETT	120	1	4.3	9.3	11.4	6.3	0.5
TRUE	1000	-	4.6	11.0	9.5	5.3	-
BARTLETT	1000	0	4.6	10.8	11.3	6.0	4.7
BARTLETT	1000	1	4.7	10.9	10.2	5.6	0.4
TRUE	5000	-	5.3	9.8	10.4	4.5	-
BARTLETT	5000	0	5.2	9.3	10.9	5.7	8.2
BARTLETT	5000	1	5.3	9.6	10.8	5.6	0.4

NOTES: The 5%(95%) and 10%(90%) columns report the frequency the χ^2-statistic is less(higher) than the lower(upper) 5% and 10% critical value. The value of b indicates the order of prewhitening. For definitions of spectral estimators see Table 1. HP refers to Hodrick-Prescott detrending and Δ refers to first-differencing. AVE(ξ_T) is the Monte Carlo mean of the bandwidth parameter.

TABLE 6b: COVERAGE PROBABILITIES FOR WALD TEST EXAMPLE: EXPERIMENT 1

$\rho = 0.5$

HP FILTERED DATA

SPECTRAL ESTIMATOR	T	b	5%	10%	90%	95%	AVE(ξ_T)
TRUE	120	-	4.6	8.2	9.9	4.5	-
BARTLETT	120	0	3.5	6.9	30.4	23.5	15.7
NW	120	0	3.5	8.0	27.8	20.3	5.79
BARTLETT	120	1	5.7	10.3	19.1	13.2	6.0
NW	120	1	5.1	10.9	22.8	17.5	15.4
BARTLETT	120	2	3.1	5.9	30.6	24.8	1.86
NW	120	2	3.4	7.6	31.9	23.7	3.14
TRUE	1000	-	4.4	9.6	10.6	5.0	-
BARTLETT	1000	0	4.2	8.5	15.7	10.5	34.7
NW	1000	0	4.6	8.6	17.6	11.1	13.6
BARTLETT	1000	1	4.9	9.8	10.8	7.0	12.8
NW	1000	1	5.4	9.9	14.1	8.6	96.0
BARTLETT	1000	2	3.3	6.5	23.8	16.1	4.01
NW	1000	2	4.6	8.3	18.8	12.1	13.2
TRUE	5000	-	5.5	9.6	10.1	4.4	-
BARTLETT	5000	0	4.9	9.4	12.7	6.4	59.9
NW	5000	0	3.4	7.8	12.2	7.2	29.9
BARTLETT	5000	1	5.5	9.9	9.0	4.2	22.1
NW	5000	1	3.9	8.5	9.6	5.4	116.6
BARTLETT	5000	2	3.9	7.8	22.3	14.9	6.97
NW	5000	2	3.4	7.7	12.4	7.3	33.5

FIRST DIFFERENCED DATA

SPECTRAL ESTIMATOR	T	b	5%	10%	90%	95%	AVE(ξ_T)
TRUE	120	-	4.3	9.9	10.1	5.0	-
BARTLETT	120	0	4.3	9.1	14.3	8.4	6.1
BARTLETT	120	1	4.8	10.0	11.5	7.5	1.0
TRUE	1000	-	5.6	12.2	9.2	5.6	-
BARTLETT	1000	0	5.0	11.9	12.5	6.6	13.2
BARTLETT	1000	1	5.8	12.6	9.3	5.0	1.6
TRUE	5000	-	5.4	9.6	10.0	5.1	-
BARTLETT	5000	0	5.4	9.6	11.1	6.7	22.7
BARTLETT	5000	1	5.7	10.1	9.4	4.2	2.6

For notes see Table 6a.

TABLE 7a: COVERAGE PROBABILITIES FOR WALD TEST EXAMPLE: EXPERIMENT 2

$\rho = 0.1$ **HP FILTERED DATA**

SPECTRAL ESTIMATOR	T	b	5%	10%	90%	95%	AVE(ξ_T)
TRUE	120	-	3.3	9.1	11.7	5.1	-
BARTLETT	120	0	3.3	7.6	21.3	14.4	10.4
NW	120	0	3.6	8.9	16.9	9.7	4.3
BARTLETT	120	1	6.5	13.7	7.0	4.8	2.8
NW	120	1	3.9	9.8	20.4	13.6	23.8
BARTLETT	120	2	3.4	8.4	20.0	14.1	.70
NW	120	2	4.3	8.2	21.3	14.8	3.8
TRUE	1000	-	5.2	9.4	9.2	5.0	-
BARTLETT	1000	0	4.8	8.6	14.2	8.2	23.3
NW	1000	0	4.8	8.8	13.6	7.8	8.7
BARTLETT	1000	1	6.9	12.5	3.9	2.2	6.2
NW	1000	1	5.0	9.0	13.4	7.8	78.1
BARTLETT	1000	2	4.8	8.8	13.3	7.8	.80
NW	1000	2	5.9	10.5	14.2	8.5	8.6
TRUE	5000	-	4.8	10.0	9.4	5.0	-
BARTLETT	5000	0	4.6	9.4	10.4	6.3	40.3
NW	5000	0	4.5	9.3	11.5	6.6	23.8
BARTLETT	5000	1	5.6	11.2	6.0	2.6	10.7
NW	5000	1	4.8	9.4	9.6	5.2	103.4
BARTLETT	5000	2	4.3	9.4	11.2	6.7	1.1
NW	5000	2	4.6	8.6	11.0	5.4	10.9

For notes see Table 6a.

TABLE 7b: COVERAGE PROBABILITIES FOR WALD TEST EXAMPLE: EXPERIMENT 2

$\rho = 0.5$ **HP FILTERED DATA**

SPECTRAL ESTIMATOR	T	b	5%	10%	90%	95%	AVE(ξ_T)
TRUE	120	-	4.2	9.3	10.7	5.9	-
BARTLETT	120	0	3.4	7.9	16.3	10.5	15.7
NW	120	0	4.8	10.5	8.3	3.9	5.1
BARTLETT	120	1	8.5	19.1	3.1	1.6	5.7
NW	120	1	5.9	13.6	12.1	8.7	28.2
BARTLETT	120	2	3.7	8.8	14.7	8.8	1.1
NW	120	2	5.1	10.0	16.2	8.5	3.8
TRUE	1000	-	4.6	9.1	10.4	5.3	-
BARTLETT	1000	0	4.1	8.6	13.3	7.3	34.7
NW	1000	0	4.7	9.0	10.6	5.3	11.6
BARTLETT	1000	1	7.5	13.4	1.9	0.7	12.4
NW	1000	1	3.6	7.9	18.3	12.8	280.1
BARTLETT	1000	2	4.6	9.0	11.2	5.7	1.8
NW	1000	2	3.9	9.1	10.3	5.8	5.8
TRUE	5000	-	6.5	10.7	10.5	5.1	-
BARTLETT	5000	0	6.6	10.7	12.0	5.3	59.9
NW	5000	0	6.6	10.8	11.6	5.5	27.4
BARTLETT	5000	1	9.1	14.8	2.5	0.6	21.4
NW	5000	1	6.1	10.2	14.7	9.1	680.9
BARTLETT	5000	2	6.7	11.2	9.8	4.5	3.2
NW	5000	2	4.5	8.9	9.0	3.3	10.2

For notes see Table 6a.