A Similarity-Based Approach to Time-Varying Coefficient Nonstationary Autoregression

By

Offer Lieberman[1]

University of Haifa

August 19, 2010

[1]Support from Israel Science Foundation grant No. 355/06 is gratefully acknowledged. I am indebted to Dan Lederman for research assistance and to Gabi Gayer and Itzhak Gilboa for numerous helpful comments. Address for correspondence: Department of Economics, University of Haifa, Haifa 31905, Israel. E-mail: offerl@econ.haifa.ac.il

Key words and phrases: Autoregression; Consistency; Dual Stocks; Nonstationarity; Similarity; Time-varying.

Abstract

We suggest in this paper a similarity based approach to time varying coefficient nonstationary autoregression. In a given sample, the model can display characteristics consistent with stationary, unit root and explosive behavior, depending on the similarity between the dependent variable and its past values. We establish consistency of the quasi maximum likelihood estimator of the model, with a general norming factor. Asymptotic score-based hypothesis tests are derived. The model is applied to a data set comprised of dual stocks traded in NASDAQ and the Tokyo Stock Exchange.

1 Introduction

We consider the stochastic process

$$\begin{aligned} Y_1 &= \mu + \varepsilon_1, \\ Y_t &= \mu + \beta_t\left(x_t, x_{t-1}; w\right) Y_{t-1} + \varepsilon_t, t = 2, ..., n, \end{aligned} \tag{1}$$

where $\mu \neq 0$, $x_t = (X_{1t}, ..., X_{mt})$ is an m-vector of characteristics of Y_t, w is an m-dimensional vector of unknown parameters, assumed to lie in a subset of R^m, $\beta_t\left(x_t, x_{t-1}; w\right)$ is a real valued, non-negative and nonstochastic function, parametrized by w, and $\{\varepsilon_t\}$ is a sequence of iid random variables with zero mean and variance σ^2. The function $\beta_t\left(x_t, x_{t-1}; w\right)$ measures the similarity between Y_t and Y_{t-1} through their characteristics. When $\beta_t\left(x_t, x_{t-1}; w\right) = 1$ for all t, the model collapses to the well known special case

$$Y_t = \mu + Y_{t-1} + \varepsilon_t, t = 2, ..., n, \tag{2}$$

i.e., a random walk with a drift. Otherwise, the coefficient on Y_{t-1} can be equal to-, greater than-, or less than unity. In other words, for a given t, the model can behave in a stationary-, unit-root- or explosive manner.

To fix ideas, consider the special case where Y_t is a (positive) price series, $m = 1$, $w > 0$ and

$$\beta_t\left(x_t, x_{t-1}; w\right) = \exp\left(w \Delta X_t\right). \tag{3}$$

If at time t, $X_t = X_{t-1}$, for this t the model is consistent with a unit root with a drift. However, if at time t, $\Delta X_t > 0$, Y_{t-1} has a coefficient consistent with explosive behavior, so that, for this t,

$$\Delta Y_t > Y_t - \beta_t\left(x_t, x_{t-1}; w\right) Y_{t-1} = \mu + \varepsilon_t.$$

For this t then, the return is expected to lie above the drift. The converse holds when, in the same setting, $\Delta X_t \leq 0$. In this case, Y_{t-1} has a coefficient consistent with a stationary behavior and ΔY_t is expected to lie under the drift, for this t. The point is that while the average Y_{t-1}-coefficient value may be close to unity, at each time period perturbations from this value are allowed, depending on the similarity between Y_t and Y_{t-1}, through their corresponding characteristics.

A similar reasoning leads to an extension of (3), viz., the model (1) with

$$\beta_t\left(x_t, x_{t-1}; w\right) = \exp\left(\Delta X_t\left(w_1 \mathbf{1}\left\{\Delta X_t > 0\right\} + w_2 \mathbf{1}\left\{\Delta X_t \leq 0\right\}\right)\right). \tag{4}$$

Here, the period-t correction for the unit-coefficient depends on the sign of ΔX_t and in some cases it might be sensible to postulate that $w_1 > w_2 > 0$, as we discuss in the empirical application section. While for many economic time series the unit-root hypothesis cannot be rejected (see, for instance, Hayashi (2000) and the references therein), similarity-based short-run perturbations from the unit-root may lead to drastic improvements in goodness of fit, as we show in Section 6.

The model (1) is a nonlinear first-order autoregression with a time-varying coefficient. The traditional specification for such a model has the form

$$Y_t = \mu + \alpha_t Y_{t-1} + \varepsilon_t, \tag{5}$$

where α_t is a *random variable*, which needs to satisfy certain regularity conditions. These conditions are usually set such that the process (5) is at least weakly stationary. The literature on time-varying coefficient autoregressions includes, among many, Nicholls and Quinn (1980, 1981, 1982), Chen and Tsay (1993), Dahlhaus *et. al.* (1999), and Lundbergh *et. al.* (2003). There is also a large body of literature on varying-coefficient models which are not autoregressions, such as the work by Hastie and Tibshirani (1993), Fan and Zhang (1999), Cai *et. al.* (2000), Cai (2002, 2007), Cai and Li (2008), Cai *et. al.* (2009), Xiao (2009), Hoderlein (2010), as well as models for nonlinear cointegrating regressions, covered by Park and Phillips (2001) and Wang and Phillips (2009a, 2009b).

It should be emphasized that none of the aforementioned references provides distribution theory for parameter estimation and hypothesis testing in model (5) when α_t is allowed to be less than-, equal to-, or greater than unity, for each t. Structurally, the key idea in our model (1) is that the coefficient on Y_{t-1} is allowed to vary with t not in a random way, but rather with the similarity between Y_t and Y_{t-1}, in line with the concept developed in the empirical similarity model of Gilboa *et. al.* (2006)

Similarity-based models were introduced to Economics by Gilboa *et. al.* (2006), applied in the context of real estate prices by Gilboa *et. al.* (2007), and suggested as an approach for prediction by Gilboa *et. al.* (2010). For the model

$$Y_t = \frac{\sum_{i<t} s_w (x_t, x_i) Y_i}{\sum_{i<t} s_w (x_t, x_i)} + \varepsilon_t, t = 2, ..., n, \tag{6}$$

where $s_w (x_t, x_i)$ is a non-negative similarity function, the asymptotic theory of estimation was established by Lieberman (2010). In (6), the sum of the coefficients on the Y_i's ($i < t$) equals one, so the model has a unit root, and the mean is set to zero. On the other hand, in (1), the coefficient on Y_{t-1}

for each t is allowed to lie in $[0, \infty)$ and the mean parameter is unknown. These differences are fundamental in the normalization rates required for the development of asymptotic theory of estimation and testing. For discussion on the rates required in the much simpler, *fixed coefficient*, nonstationary case, with or without a drift, the reader is referred to Evans and Savin (1984, eq'n (4.1)) or Phillips (1987, p 545).

For the proposed model, (1), we prove consistency of the QMLE with a general norming factor, previously unused, which includes in it as special cases, the factors required in the fixed coefficient world, in any of the stable-, unit root- or explosive cases, with or without a drift. We investigate the asymptotic behavior of the score and Hessian of the concentrated log-likelihood and use it to construct asymptotic score-based (i.e., Lagrange Multiplier) hypothesis tests for the model parameters when the process behaves as a time varying coefficient, approximately unit root process with a drift. By approximately-, it is meant that the norming factors required are in agreement with those used in the simple unit root with a drift model and by asymptotic-, it is meant that the test is based on the leading terms of the score and the Hessian. We show that these leading terms obey the second-order Bartlett identities. The results are then applied in an empirical application involving dual stocks traded in both the Tokyo Stock Exchange and in NASDAQ.

To summarize, the main contributions that this paper makes are as follows. First, the specification (1) is proposed as stemming from the axiomatically justified economic theory of similarity-based modeling. Secondly, we prove consistency of the QMLE of the parameter vector when $\beta_t(x_t, x_{t-1}; w)$ is allowed to lie in $[0, \infty)$ for each t, with a general norming factor. It should be stressed that to our knowledge, none of the varying coefficient autoregression models studied to date cover this scenario. Third, we show that the general norming factor collapses to the well known factors required in the fixed coefficient world, in any of the stable-, unit root- or explosive cases, with or without a drift. Fourth, we derive asymptotic score-based tests for hypothesis testing in the model. Finally, we apply the theory to Japanese stocks which are dually listed in the Tokyo Stock exchange and in NASDAQ.

The plan for the paper is as follows. In Section 2 we present the setup and discuss the model assumptions. Consistency is established in Section 3. Asymptotics for the concentrated score and Hessian are derived in Section 4 and used in Section 5 to construct asymptotic score-based hypothesis tests. An empirical application to dual stock data follows in Section 6 and Section 7 concludes.

2 Assumptions and Set-up

For the model (1), set

$$\beta_t = \beta_t\left(x_t, x_{t-1}; w\right).$$

In matrix form, the model is $Sy = \mu 1 + \varepsilon$, where $S = S\left(X, w\right) = I_n - C\left(X, w\right)$, I_n is the identity matrix of order n,

$$C = C\left(X, w\right) = \begin{pmatrix} 0 & 0 & \cdots & & & 0 \\ \beta_2 & 0 & 0 & \cdots & & 0 \\ 0 & \beta_3 & 0 & 0 & \cdots & 0 \\ \cdots & & \cdots & & & \cdots \\ 0 & \cdots & & 0 & \beta_n & 0 \end{pmatrix},$$

X is assumed to be nonstochastic with rows $(x_t)_{1 \le t \le n}$, $y = (Y_1, ..., Y_n)'$, $\varepsilon = (\varepsilon_1, ...\varepsilon_n)'$ and 1 is an n-vector of 1's. Note that C is nonnegative and nilpotent. The Jacobian of transformation from ε to y is unity, so the quasi-log-likelihood is

$$l_n\left(\theta\right) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{\left(Sy - \mu 1\right)'\left(Sy - \mu 1\right)}{2\sigma^2}.$$

We denote the quasi maximum likelihood estimators (QMLE's) of μ and w by $\hat{\mu}_n$ and $\hat{w}_n$, respectively and note that $\hat{\mu}_n = 1'S\left(X, \hat{w}_n\right)y/n$. Hence, the concentrated quasi-log-likelihood function is equal to

$$l_n^c\left(\theta\right) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{y'S'MSy}{2\sigma^2}, \tag{7}$$

where $M = I_n - 11'/n = I_n - P$, say.

Let $\theta = \left(\sigma^2, w_1, \ldots, w_m\right)' = \left(\theta_1, \theta_2'\right)'$ with $\sigma^2 = \theta_1$, $\theta_2' = (w_1, \ldots, w_m)$, and denote the true values of θ and μ by θ_0 and μ_0, respectively. The parameter space is $\Theta = \Theta_1 \times \Theta_2$, where Θ_1, Θ_2 are the spaces in which σ^2 and w are assumed to lie, respectively. By K we denote a generic bounding constant, independent of n, which may vary from step to step. In the following we enlist the assumptions used for our model.

Assumption A0: $\{\varepsilon_t\}_{t=1}^n$ is a sequence of iid continuous random variables, each with a zero mean, variance σ^2 and bounded cumulants κ_r, $r \ge 3$. If $w \ne w'$, the set $\left\{X \middle| \left[C\left(X, w\right)\right]_{i,j} \ne \left[C\left(X, w'\right)\right]_{i,j}\right\}$ has a positive Lebesgue measure for all $i = 3, ..., n$ and $j < i$. The matrix X is allowed to lie in $\tilde{X}_{n,m}$, the set of all $(n \times m)$ nonstochastic, real and finite matrices.

Assumption A1**:** There exist σ_L^2, σ_H^2, w_L and w_H, such that $\sigma_0^2 \in \left[\sigma_L^2, \sigma_H^2\right]$, with $0 < \sigma_L^2 < \sigma_H^2 < \infty$ and for each $i = 1, ..., m$, $w_{i,0} \in [w_L, w_H]$, with $-\infty < w_L < w_H < \infty$. In addition, $\mu_0 \in \mathbb{R} \backslash \{0\}$.

Assumption A2: For all $1 < t \leq n$, the function $\beta_t\left(x_t, x_{t-1}; w\right)$ is nonnegative, continuous in x and in w and is twice-continuously differentiable in w.

Let $C_0 = C\left(X, w_0\right)$, so that $S_0 = I_n - C_0$ and set

$$\dot{C}_r\left(X, w\right) = \partial C\left(X, w\right) / \partial w_r \text{ and } \ddot{C}_{r,s}\left(X, w\right) = \partial^2 C\left(X, w\right) / \partial w_r \partial w_s.$$

Assumption A3: For all $1 \leq r \leq m$, $1 \leq i, j \leq n$, for all $X \in \tilde{X}_{n,m}$ and for all $w \in \Theta_2 \subset R^m$,

$$C_{i,j} \leq K C_{0i,j},$$

and

$$\left| \left[\dot{C}_r\left(X, w\right) \right]_{i,j} \right| \leq K \left[C\left(X, w\right)\right]_{i,j}.$$

Assumption A4: $\ddot{C}_{r,s}\left(X, w\right)$ is continuous at all (X, w) and for all $1 \leq r, s \leq m$, $1 \leq i, j \leq n$,

$$\left| \left[\ddot{C}_{r,s}\left(X, w\right) \right]_{i,j} \right| \leq K \left[C\left(X, w\right)\right]_{i,j}.$$

Assumptions A0–A4 are similar to Assumptions A0–A4 of Lieberman (2010), who were set on model (6), apart for the following differences. In Assumptions A2, w need not be nonnegative and Assumption A3 is less restrictive than Lieberman's (2010) Assumption A3. We do not need a condition on the third-order derivative for our work. Model (1) includes a drift term and β_t is allowed to lie in $[0, \infty)$, whereas in (6) the weights on past Y_i's sum up to unity by construction and there is no drift term. It is easy to verify that Assumptions A0-A4 hold for model (1) with β_t given by (3) and (4).

By $\|A\|_1$, $\|A\|_2$ and $\|A\|_F$ we denote the l_1-, spectral- and Frobenius norms of an $n \times n$ matrix A. That is, $\|A\|_1 = \sum_{i,j=1}^n \left|[A]_{i,j}\right|$, $\|A\|_2 = \left(\sup_{|x|=1} x'A'Ax\right)^{1/2}$ and $\|A\|_F = \left(tr\left(A'A\right)\right)^{1/2}$. We shall use the following inequalities (see, for instance, Graybill (1983)).

$$\begin{aligned} \|A\|_2 &\leq \|A\|_F \leq \sqrt{n}\,\|A\|_2, \\ 1'A1 &\leq n\,\|A\|_2, \text{ for } A > 0, \; |tr\left(AB\right)| \leq \|A\|_F \|B\|_F, \\ \|AB\|_F &\leq \|A\|_2 \|B\|_F, \; \|AB\|_F \leq \|A\|_F \|B\|_2. \end{aligned} \tag{8}$$

For the matrix S, we remark that as C is nonnegative, for all $\theta_2 \in \Theta_2$ and for all $X \in \tilde{X}_{n,m}$,

$$\left\| S^{-1} \right\|_F^2 = \sum_{i,j=1}^{n} \left[I + C + C^2 + \cdots + C^{n-1} \right]_{i,j}^2 \geq n, \tag{9}$$

with equality if and only if $C = 0$. In addition,

$$\left\| S^{-1} \right\|_F^2 \leq \left\| S^{-1\prime} S^{-1} \right\|_1. \tag{10}$$

3 Consistency

In this section we prove consistency of the QMLE with a general norming factor. As is well known, to prove consistency of the QMLE of β in the fixed coefficient model,

$$\begin{aligned} Y_1 &= \mu + \varepsilon_1, \\ Y_t &= \mu + \beta Y_{t-1} + \varepsilon_t, t = 2, ..., n, \end{aligned} \tag{11}$$

the quasi-log-likelihood must be normalized by $(1/n)$, when $|\beta_0| < 1$, by $(1/n^2)$, when $\beta_0 = 1$ and $\mu = 0$, by $(1/n^3)$, when $\beta_0 = 1$ and $\mu \neq 0$, and by β^{-2n}, when $\beta_0 > 1$. See, for instance, Evans and Savin[2] (1984, eq'n (4.1)) or Phillips (1987, p 545). In our model (1) though, for each t, β_t can be smaller than-, equal to- or greater than unity. Hence, the aforementioned normalization factors are generally invalid in our case. To overcome this problem, for a proof of consistency of $\hat{w}_n$ in the $\mu \neq 0$ case, we normalize the quasi-log-likelihood by $\left\| S_0^{-1\prime} S_0^{-1} \right\|_1^{-1}$. It turns out that this term collapses to the required rates in each of the aforementioned special cases. To justify the use of this norming factor, we first prove the following Lemma.

Lemma 1 *Let*

$$\rho_n = \frac{\left\| S_0^{-1} \right\|_F^2}{\left\| S_0^{-1\prime} S_0^{-1} \right\|_1}.$$

[2] The model (11) is different from the one investigated by Evans and Savin (1984) with regards to starting value Y_1.

For the model (11), there exists a constant c_L, independent of n, with $0 < c_L < 1$, and there exists an $\mathbb{N}$, such that $\forall n \geq \mathbb{N}$,

$$\begin{aligned} c_L &< \rho_n \leq 1, \text{ if } \beta < 1, \\ \rho_n &= \frac{3}{2n+1}, \text{ if } \beta = 1, \\ c_L &< \rho_n \leq 1, \text{ if } 1 < \beta < \infty. \end{aligned} \tag{12}$$

Proof of Lemma 1: For the model (11), the matrix S_0^{-1} is easily seen to equal $\left[S_0^{-1}\right]_{i,j} = \beta^{i-j} 1\left\{i \geq j\right\}$. Hence, for large n, there exist constants c_{L1} and K^{U1}, independent of n, with $0 < c_{L1} < K^{U1} < \infty$, such that

$$c_{L1} n < \left\|S_0^{-1}\right\|_F^2 = n \sum_{j=1}^{n-1} \left(1 - j/n\right) \beta^{2j} < K^{U1} n, \text{ if } \beta < 1,$$

$$\left\|S_0^{-1}\right\|_F^2 = n\left(n+1\right)/2, \text{ if } \beta = 1 \tag{13}$$

and

$$c_{L1} \beta^{2n} < \left\|S_0^{-1}\right\|_F^2 = \beta^{2n} \sum_{j=1}^{n} \left(j/\beta^{2j}\right) < K^{U1} \beta^{2n}, \text{ if } \beta > 1.$$

On the other hand, it is easy to see that for large n, there exist constants c_{L2} and K^{U2}, independent of n, with $0 < c_{L2} < K^{U2} < \infty$, such that

$$\begin{aligned} \frac{c_{L2} n}{1-\beta^2} &< \left\|S_0^{-1\prime} S_0^{-1}\right\|_1 < \frac{K^{U2} n}{1-\beta^2}, \text{ if } \beta < 1, \\ \left\|S_0^{-1\prime} S_0^{-1}\right\|_1 &= \frac{n\left(n+1\right)\left(2n+1\right)}{6}, \text{ if } \beta = 1, \\ c_{L2} \beta^{2n} &< \left\|S_0^{-1\prime} S_0^{-1}\right\|_1 < K^{U2} \beta^{2n}, \text{ if } \beta > 1. \end{aligned} \tag{14}$$

Finally, by the inequality (10), $\rho_n \leq 1$.■

In the stable- and explosive cases then, ρ_n is bounded from below and from above, whereas it behaves as n^{-1} in the $\beta = 1$ case. We remark that ρ_n is at most $O\left(1\right)$ for the more general varying coefficient model (1) simply by virtue of the inequality (10). Moreover, it appears from (14) that normalizing the quasi-log-likelihood by $\left\|S_0^{-1\prime} S_0^{-1}\right\|_1^{-1}$, is consistent with a $1/n$ - rate required in the $\beta < 1$ case, with a $1/n^3$ - rate in the $\beta = 1$ case, and with a rate of $1/\beta^{2n}$ in the $\beta > 1$ case. For the $\beta < 1$ and $\beta > 1$ cases, these are also the rates required when $\mu = 0$. For the $\beta = 1$ and $\mu = 0$ case, the required rate is $1/n^2$.

We are now ready to pursue the result of consistency.

Theorem 2 *Under Assumptions A0-A2,* $\hat{\theta}_n \to_p \theta$.

Proof of Theorem 2: We note that

$$\begin{aligned} l_n^c(\theta_0) - l_n^c(\theta) &= \left(l_n^c\left(\sigma_0^2, \theta_{20}'\right) - l_n^c\left(\sigma^2, \theta_{20}'\right)\right) & (15) \\ &\quad + \left(l_n^c\left(\sigma^2, \theta_{20}'\right) - l_n^c\left(\sigma^2, \theta_2'\right)\right) & (16) \end{aligned}$$

and that in view of (7), equation (16) is equal to

$$\frac{\sigma_0^2}{\sigma^2}\left(l_n^c\left(\sigma_0^2, \theta_{20}'\right) - l_n^c\left(\sigma_0^2, \theta_2'\right)\right). \tag{17}$$

For any $\delta_1 > 0$, denote by $B_{\delta_1}(\theta_0)$ the ball $\{\theta \in \Theta : ||\theta - \theta_0|| \leq \delta_1\}$ and by $B_{\delta_1}^c(\theta_0)$ the complement of $B_{\delta_1}(\theta_0)$ in Θ. Using (15)-(17), for the proof of consistency it is sufficient to show that $\forall \delta_1 > 0$,

$$\liminf_{n\to\infty} \inf_{B_{\delta_1}^c(\theta_0)} n^{-1}\left(l_n^c\left(\sigma_0^2, \theta_{20}'\right) - l_n^c\left(\sigma^2, \theta_{20}'\right)\right) \tag{18}$$

and

$$\liminf_{n\to\infty} \inf_{B_{\delta_1}^c(\theta_0)} \left\|S_0^{-1\prime} S_0^{-1}\right\|_1^{-1} \left(l_n^c\left(\sigma_0^2, \theta_{20}'\right) - l_n^c\left(\sigma_0^2, \theta_2'\right)\right) \tag{19}$$

are strictly positive in probability. See Wu (1981).

Now, using the results on the cumulants of quadratic forms in non-normal variables (e.g., Lieberman (1997)) and the fact that $tr(M) = n-1$,

$$E_{\theta_0}\left(n^{-1} l_n^c\left(\sigma^2, \theta_{20}'\right)\right) = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{(n-1)\sigma_0^2}{2\sigma^2 n}$$

and

$$Var_{\theta_0}\left(n^{-1} l_n^c\left(\sigma^2, \theta_{20}'\right)\right) = \frac{\sigma_0^4(n-1)}{2\sigma^4 n^2} + \frac{\kappa_4(n-1)^2}{4\sigma^4 n^3}.$$

Hence,

$$n^{-1}\left(l_n^c\left(\sigma_0^2, \theta_{20}\right) - l_n^c\left(\sigma^2, \theta_{20}'\right)\right) \to_p \frac{1}{2}\left(\frac{\sigma_0^2}{\sigma^2} - 1 - \log\left(\frac{\sigma_0^2}{\sigma^2}\right)\right) \geq 0,$$

with equality iff $\sigma^2 = \sigma_0^2$.

To establish (19), set

$$G = S - S_0 = C_0 - C.$$

Then,

$$\begin{aligned}
\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{-1}\left(l_n^c\left(\sigma_0^2,\theta_{20}'\right)-l_n^c\left(\sigma_0^2,\theta_2'\right)\right) &= \frac{1}{2\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1}\left(y'S'MSy-y'S_0MS_0y\right)\\
&= \frac{1}{2\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1}y'S_0'\left(S_0^{-1\prime}G'M+MGS_0^{-1}\right)S_0y\\
&\quad+\frac{1}{2\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1}y'G'MGy,\\
&= Q_{1n}+Q_{2n}, \qquad (20)
\end{aligned}$$

say. Since $M1=0$,

$$y'S_0'\left(S_0^{-1\prime}G'M+MGS_0^{-1}\right)S_0y=\left(S_0y-\mu_01\right)'\left(S_0^{-1\prime}G'M+MGS_0^{-1}\right)\left(S_0y-\mu_01\right).$$

As C and C_0 are nilpotent, G is nilpotent and as S is lower triangular, so is S^{-1}. Hence, GS^{-1} is lower triangular, implying that,

$$\begin{aligned}
\left|E_{\theta_0}\left(Q_{1n}\right)\right| &= \frac{\left|tr\left(MGS_0^{-1}\right)\right|}{\left\|S_0^{-1\prime}S_0^{-1}\right\|_1}\\
&\leq \frac{\|P\|_F\left\|GS_0^{-1}\right\|_F}{\left\|S_0^{-1\prime}S_0^{-1}\right\|_1}\\
&\leq K\rho_n\left\|S_0^{-1}\right\|_F^{-1}, \qquad (21)
\end{aligned}$$

by Assumption A3 and because $\left\|C_0S_0^{-1}\right\|_F\leq\left\|S_0^{-1}\right\|_F$. Further,

$$\begin{aligned}
Var_{\theta_0}\left(Q_{1n}\right) &= \frac{1}{2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^2}tr\left(S_0^{-1\prime}G'M+MGS_0^{-1}\right)^2\\
&\quad+\frac{1}{4\sigma_0^4\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^2}\kappa_4\sum_{i=1}^n\left[S_0^{-1\prime}G'M+MGS_0^{-1}\right]_{i,i}^2. \qquad (22)
\end{aligned}$$

The first term in (22) is equal to

$$\frac{1}{\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^2}tr\left(\left(MGS_0^{-1}\right)^2+S_0^{-1\prime}G'MGS_0^{-1}\right).$$

We have

$$\begin{aligned}
tr\left(\left(MGS_0^{-1}\right)^2\right) &= tr\left(\left(GS_0^{-1}\right)^2-GS_0^{-1}PGS_0^{-1}-P\left(GS_0^{-1}\right)^2+\left(PGS_0^{-1}\right)^2\right)\\
&\leq 2tr\left(P\left(GS_0^{-1}\right)^2\right)+\left(tr\left(PGS_0^{-1}\right)\right)^2\\
&\leq K\left\|S_0^{-1}\right\|_F^2,
\end{aligned}$$

Also,

$$\begin{aligned}\left|tr\left(S_0^{-1\prime}G'MGS_0^{-1}\right)\right| &\leq \left\|MGS_0^{-1}\right\|_F\left\|S_0^{-1\prime}G'\right\|_F\\ &\leq \|M\|_2\left\|GS_0^{-1}\right\|_F^2\\ &\leq K\left\|S_0^{-1}\right\|_F^2.\end{aligned}$$

The second part in (22) involves

$$\begin{aligned}\sum_{i=1}^n\left[S_0^{-1\prime}G'M+MGS_0^{-1}\right]_{i,i}^2 &\leq 4tr\left(PGS_0^{-1}\right)^2\\ &\leq K\left\|S_0^{-1}\right\|_F^2.\end{aligned}$$

It follows that $Var_{\theta_0}\left(Q_{1n}\right)=O\left(\rho_n^2\left\|S_0^{-1}\right\|_F^{-2}\right)$ and together with (21) and Chebyshev's inequality, we deduce that

$$Q_{1n}=O_p\left(\rho_n\left\|S_0^{-1}\right\|_F^{-1}\right). \tag{23}$$

Next, we consider Q_{2n}, defined in (20). We have

$$\begin{aligned}y'G'MGy &= \varepsilon'S_0^{-1\prime}G'MGS_0^{-1}\varepsilon+2\mu_0 1'S_0^{-1\prime}G'MGS_0^{-1}\varepsilon+\mu_0^2 1'S_0^{-1\prime}G'MGS_0^{-1}1\\ &= q_{21n}+q_{22n}+q_{23n}, \qquad (24)\end{aligned}$$

say. Now,

$$\begin{aligned}E_{\theta_0}\left(q_{21n}\right) &= \sigma_0^2 tr\left(S_0^{-1\prime}G'MGS_0^{-1}\right)\\ &= \sigma_0^2\left\|MGS_0^{-1}\right\|_F^2\\ &\leq K\left\|S_0^{-1}\right\|_F^2\end{aligned}$$

and

$$\begin{aligned}Var_{\theta_0}\left(q_{21n}\right) &= 2\sigma_0^4 tr\left(S_0^{-1\prime}G'MGS_0^{-1}\right)^2+\kappa_4\sum_{i=1}^n\left[S_0^{-1\prime}G'MGS_0^{-1}\right]_{i,i}^2\\ &\leq K\left\|S_0^{-1}\right\|_F^4,\end{aligned}$$

so that $q_{21n}=O_p\left(\left\|S_0^{-1}\right\|_F^2\right)$. The second term in (24) has

$$E_{\theta_0}\left(q_{22n}\right)=0$$

and

$$
\begin{aligned}
Var_{\theta_0}\left(q_{22n}\right) &= 4\mu_0^2\sigma_0^2 1'\left(S_0^{-1\prime}G'MGS_0^{-1}\right)^2 1 \\
&\leq K\left(1'\left(S_0^{-1\prime}G'GS_0^{-1}\right)^2 1 + 1'\left(S_0^{-1\prime}G'PGS_0^{-1}\right)^2 1\right). \quad (25)
\end{aligned}
$$

Using the inequality $x'Ax \leq x'x\left\|A\right\|_2$, for $A > 0$, the first term in (25) is bounded by

$$
\begin{aligned}
K1'\left(S_0^{-1\prime}S_0^{-1}\right)^2 1 &\leq K1'S_0^{-1\prime}S_0^{-1}1\left\|S_0S_0^{-1\prime}\right\|_2 \\
&\leq K\left\|S_0^{-1\prime}S_0^{-1}\right\|_1\left\|S_0^{-1}\right\|_F^2 = K\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^2\rho_n.
\end{aligned}
$$

The second term equals

$$
\begin{aligned}
& Ktr\left(PGS_0^{-1}S_0^{-1\prime}G'PGS_0^{-1}11'S_0^{-1\prime}G'\right) \\
\leq\ & K\left\|P\right\|_F\left\|GS_0^{-1}S_0^{-1\prime}G'PGS_0^{-1}11'S_0^{-1\prime}G'\right\|_F \\
\leq\ & K\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^2\rho_n.
\end{aligned}
$$

This implies that $q_{22n} = O_p\left(\sqrt{\rho_n}\left\|S_0^{-1\prime}S_0^{-1}\right\|_1\right)$. The last term in (24) is

$$
\mu_0^2 1'S_0^{-1\prime}G'MGS_0^{-1}1 \leq K\left\|S_0^{-1\prime}S_0^{-1}\right\|_1.
$$

It follows that

$$
Q_{2n} = O_p(1) + O_p\left(\sqrt{\rho_n}\right) + O_p(\rho_n),
$$

where $O_p(\rho_n)$ is at most $O_p(1)$, by (10). Given (9), (10) and (23), Q_{1n} is negligible. Under Assumption A0, $S_0^{-1\prime}G'MGS_0^{-1}$ is positive semidefinite and non-null and y is a continuous random variable. It follows that Q_{2n} is strictly positive in probability uniformly in $B_{\delta_1}^c(\theta_0)$ and the proof of consistency is completed.■

We emphasize that while the proof is stated under Assumptions A0-A2, including Assumption A1, under which $\mu_0 \neq 0$, the norming factor is also correct for the $\mu = 0$ and $\beta < 1$ or $\beta > 1$ cases. For the $\beta = 1$ and $\mu = 0$ case, the required rate is n^{-2} (see, for instance, Phillips (1987)).

4 Asymptotics for the Score and Hessian

In this section we investigate the asymptotic behavior of the concentrated score vector and Hessian matrix and use the results in the next section to construct asymptotic score–based (Lagrange Multiplier) hypothesis tests for the model. Let

$$D_n = \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\left\|S_0^{-1\prime} S_0^{-1}\right\|_1^{1/2}} & & \\ \cdots & & \cdots & \\ 0 & & & \frac{1}{\left\|S_0^{-1\prime} S_0^{-1}\right\|_1^{1/2}} \end{pmatrix}$$

and define the normalized concentrated score by

$$z_n(\theta) = D_n \frac{\partial l_n^c(\theta)}{\partial \theta}.$$

We shall concentrate on the case $\rho_n = o(1)$. The stronger condition $\rho_n = O\left(n^{-1}\right)$, given in (12) and holding for a unit root process-type, is no doubt prevalent in much economic data. See Hayashi (2000). The following lemma gives the asymptotic behavior of $z_n(\theta)$.

Lemma 3 *Under Assumptions A0-A3 and under the condition $\rho_n = o(1)$,*

1. $z_{n1}(\theta_0) \to_d N\left(0, \frac{1}{2\sigma_0^4} + \frac{\kappa_4}{4\sigma_0^8}\right)$.
2. For $2 \leq r \leq m+1$,

   $$z_{nr}(\theta) = z_{nr}^L(\theta_0) + o_p(1),$$

   where

   $$z_{nr}^L(\theta_0) = \frac{-\mu_0 1' S_0^{-1\prime} \dot{S}_{0r}' M\varepsilon}{\sigma_0^2 \left\|S_0^{-1\prime} S_0^{-1}\right\|_1^{1/2}} = O_p(1). \tag{26}$$

3. $z_{n1}(\theta_0)$ is asymptotically (as $n \to \infty$) independent of $z_{nr}(\theta_0)$, $2 \leq r \leq m+1$.

Proof of Lemma 3: The concentrated score with respect to σ^2 is given by

$$z_{n1}(\theta_0) = -\frac{\sqrt{n}}{2\sigma_0^2} + \frac{y' S_0' M S_0 y}{2\sigma_0^4 \sqrt{n}}. \tag{27}$$

Because $tr(M) = n-1$, we have

$$E_{\theta_0}(z_{n1}(\theta_0)) = -\frac{1}{2\sigma_0^4 \sqrt{n}}$$

and

$$\begin{aligned} Var_{\theta_0}\left(z_{n1}\left(\theta_0\right)\right) &= \frac{2\sigma_0^4\left(n-1\right)+\kappa_4 n\left(1-n^{-1}\right)^2}{4\sigma_0^8 n} \\ &\to_{n\to\infty} \frac{1}{2\sigma_0^4}+\frac{\kappa_4}{4\sigma_0^8}. \end{aligned}$$

Higher order cumulants of $z_{n1}\left(\theta_0\right)$ tend to zero, so part 1 of the lemma is established. Further, for $r=2,...,m+1$,

$$\begin{aligned} z_{nr}\left(\theta_0\right) &= -\frac{y'S_0'\left(M\dot{S}_{0r}S_0^{-1}+S_0^{-1\prime}\dot{S}_{0r}'M\right)S_0 y}{2\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{1/2}} \qquad (28) \\ &= -\frac{1}{2\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{1/2}}(\varepsilon'\left(M\dot{S}_{0r}S_0^{-1}+S_0^{-1\prime}\dot{S}_{0r}'M\right)\varepsilon \\ &\quad +2\mu_0 1'S_0^{-1\prime}\dot{S}_{0r}'M\varepsilon), \end{aligned}$$

because $M1=0$. Now, as $\dot{S}_{0r}S_0^{-1}$ is lower triangular,

$$\begin{aligned} \left|E_{\theta_0}\left(\varepsilon'\left(M\dot{S}_{0r}S_0^{-1}+S_0^{-1\prime}\dot{S}_{0r}'M\right)\varepsilon\right)\right| &= \sigma_0^2\left|tr\left(M\dot{S}_{0r}S_0^{-1}+S_0^{-1\prime}\dot{S}_{0r}'M\right)\right| \\ &= 2\sigma_0^2\left|tr\left(P\dot{S}_{0r}S_0^{-1}\right)\right| \\ &\leq K\left\|S_0^{-1}\right\|_F, \end{aligned}$$

under Assumption A3. In addition,

$$\begin{aligned} Var_{\theta_0}\left(\varepsilon'\left(M\dot{S}_{0r}S_0^{-1}+S_0^{-1\prime}\dot{S}_{0r}'M\right)\varepsilon\right) &= 2\sigma_0^4 tr\left(M\dot{S}_{0r}S_0^{-1}+S_0^{-1\prime}\dot{S}_{0r}'M\right)^2 \\ &\quad +\kappa_4\sum_{i=1}^{n}\left[M\dot{S}_{0r}S_0^{-1}+S_0^{-1\prime}\dot{S}_{0r}'M\right]_{i,i}^2 \\ &\leq 4\sigma_0^4 tr\left(\left(M\dot{S}_{0r}S_0^{-1}\right)^2+S_0^{-1\prime}\dot{S}_{0r}'M\dot{S}_{0r}S_0^{-1}\right) \\ &\quad +\kappa_4\left(tr\left(M\dot{S}_{0r}S_0^{-1}+S_0^{-1\prime}\dot{S}_{0r}'M\right)\right)^2 \\ &\leq K\left\|S_0^{-1}\right\|_F^2. \end{aligned}$$

Thus,

$$\varepsilon'\left(M\dot{S}_{0r}S_0^{-1}+S_0^{-1\prime}\dot{S}_{0r}'M\right)\varepsilon=O_p\left(\left\|S_0^{-1}\right\|_F\right). \qquad (29)$$

Next,

$$E_{\theta_0}\left(1'S_0^{-1\prime}\dot{S}_{0r}'M\varepsilon\right)=0$$

and

$$Var_{\theta_0}\left(1'S_0^{-1\prime}\dot{S}_{0r}'M\varepsilon\right) = \sigma_0^2 1'S_0^{-1\prime}\dot{S}_{0r}'M\dot{S}_{0r}S_0^{-1}1$$
$$\leq K\left\|S_0^{-1\prime}S_0^{-1}\right\|_1,$$

under Assumption A3. Hence,

$$1'S_0^{-1\prime}\dot{S}_{0r}'M\varepsilon = O_p\left(\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{1/2}\right). \tag{30}$$

It follows from (28), (29) and (30) that under the condition $\rho_n = o(1)$,

$$z_{nr}(\theta_0) = -\frac{\mu_0 1'S_0^{-1\prime}\dot{S}_{0r}'M\varepsilon}{\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{1/2}} + o_p(1), \tag{31}$$

and the leading term of (31) is $O_p(1)$, as required. Finally, because $M1 = 0$,

$$Cov_{\theta_0}\left(\varepsilon' M\varepsilon, 1'S_0^{-1\prime}\dot{S}_{0r}'M\varepsilon\right) = 0,$$

implying that for $r = 2, ..., m+1$, $z_{n1}(\theta_0)$ and $z_{nr}(\theta_0)$ are asymptotically uncorrelated. This completes the proof of the lemma.■

We remark that if ε is a normal random vector and $\rho_n = o(1)$, then $z_{nr}(\theta_0)$ is asymptotically normal. However, if ε is non-normal, then all we can say is that the leading terms of $z_{nr}(\theta_0)$ can be expressed as $\sum_{i=1}^n h_{in}\varepsilon_i$ with non-stochastic weights h_{in} satisfying $\sum_{i=1}^n h_{in}^2 = O(1)$. If, in addition, $h_{in} = O\left(n^{-1}\right)$ for all $i = 1, ..., n$, then an Edgeworth-type expansion for the distribution of $z_{nr}(\theta_0)$ is easily derivable and $z_{nr}(\theta_0)$ will be asymptotically normal. However, this requirement cannot be ensured in general without additional structure on the model. For instance, if $h_{1n} = 1$ and $h_{jn} = 0$, $j = 2, ..., n$, then $z_{nr}(\theta_0) = -\left(\mu_0/\sigma_0^2\right)\varepsilon_1 + o_p(1)$, which is not asymptotically normal.

Let

$$H_n(\theta) = D_n \frac{\partial^2 l_n^c(\theta)}{\partial\theta\partial\theta'} D_n.$$

The next Lemma provides the asymptotic behavior of H_n.

Lemma 4 *Under Assumptions A0-A4 and the condition $\rho_n = o(1)$, for $2 \leq r, s \leq m+1$,*

$$H_{nr,s}(\theta_0) = H_{nr,s}^L(\theta_0) + o_p(1) \tag{32}$$

where

$$H_{nr,s}^L(\theta_0) = \frac{\mu_0^2 1'S_0^{-1\prime}\dot{S}_{0r}'M\dot{S}_{0s}S_0^{-1}1}{\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1} = O(1). \tag{33}$$

The proof of Lemma 4 is similar to that of Lemma 3 and only the main ingredients will be given here.

Proof of Lemma 4: For $2 \leq r, s \leq m+1$,

$$
\begin{aligned}
H_{nr,s}(\theta_0) &= -\frac{y'\left(\ddot{S}'_{0r,s} M S_0 + \dot{S}'_{0r} M \dot{S}_{0s} + \dot{S}'_{0s} M \dot{S}_{0r} + S'_0 M \ddot{S}_{0r,s}\right) y}{2\sigma_0^2 \left\| S_0^{-1\prime} S_0^{-1} \right\|_1} \\
&= -\frac{1}{2\sigma_0^2 \left\| S_0^{-1\prime} S_0^{-1} \right\|_1} \{ \varepsilon' \left(S_0^{-1\prime} \ddot{S}'_{0r,s} M + M \ddot{S}_{0r,s} S_0^{-1} \right) \varepsilon \qquad (34) \\
&\quad + 2\mu_0 1' S_0^{-1\prime} \ddot{S}'_{0r,s} M \varepsilon + 2\varepsilon' S_0^{-1\prime} \dot{S}'_{0s} M \dot{S}_{0r} S_0^{-1} \varepsilon \\
&\quad + 4\mu_0 1' S_0^{-1\prime} \dot{S}'_{0s} M \dot{S}_{0r} S_0^{-1} \varepsilon + 2\mu_0^2 1' S_0^{-1\prime} \dot{S}'_{0s} M \dot{S}_{0r} S_0^{-1} 1 \}.
\end{aligned}
$$

Under the condition $\rho_n = o(1)$, the dominant term in (34) is seen to be

$$
1' S_0^{-1\prime} \dot{S}'_{0s} M \dot{S}_{0r} S_0^{-1} 1 = O(\left\| S_0^{-1\prime} S_0^{-1} \right\|_1),
$$

with remaining terms being $o_p(1)$. The Lemma is thus proven.■

The following corollary is obvious from (31) and (32)

Corollary 5 *The second order Bartlett identity holds for the leading terms of (31) and (32). That is,*

$$
E_{\theta_0}\left(z_{nr}^L(\theta_0)\, z_{ns}^L(\theta_0)\right) + E_{\theta_0}\left(H_{nr,s}^L(\theta_0)\right) = 0. \qquad (35)
$$

5 Asymptotic Score-Based Hypothesis Tests

We can use the results of the previous section to form asymptotic score-based hypothesis tests of the form $H_0 : \theta = \theta_0$ under the condition $\rho_n = o(1)$. By asymptotic- it is meant that in the construction of the tests, only the leading terms of the score and of the Hessian are used. For $r = 2, ..., m+1$, let

$$
v'_{0r} = 1' S_0^{-1\prime} \dot{S}'_{0r} M
$$

and

$$
V'_0 = \begin{pmatrix} v'_{02} \\ \cdot \\ \cdot \\ \cdot \\ v'_{0,m+1} \end{pmatrix} = \begin{pmatrix} 1' S_0^{-1\prime} \dot{S}'_{02} M \\ \cdot \\ \cdot \\ \cdot \\ 1' S_0^{-1\prime} \dot{S}'_{0m+1} M \end{pmatrix}, \qquad (36)
$$

so that V_0' is $m \times n$. Assuming that $V_0'V_0$ is full rank, which will hold if the elements of V_0 are linearly independent, and in view of (31) and (36), we define the asymptotic score-based test as

$$T_n = \frac{y'\hat{S}_0'\hat{V}_0\left(\hat{V}_0'\hat{V}_0\right)^{-1}\hat{V}_0'\hat{S}_0 y}{\hat{\sigma}_0^2}, \tag{37}$$

where $\hat{V}_0$ and $\hat{S}_0$ are V_0 and S_0 evaluated at the QMLE's under H_0 and $\hat{\sigma}_0^2 = y'\hat{S}_0'M\hat{S}_0 y/n$. Note that since $M1 = 0$, $V_0'S_0 y = V_0'\varepsilon$. Using Theorem 1, if $\varepsilon \sim \left(0, \sigma_0^2\right)$, $T_n \overset{a}{\sim} \chi^2(m)$, under H_0. In the scalar case, when the hypothesis is of the form $H_0 : \theta_r = \theta_{0r}$, (37) reduces to

$$T_n = \frac{\left(\hat{v}_{0r}'\hat{S}_0 y\right)^2}{\hat{\sigma}_0^2 \hat{v}_{0r}'\hat{v}_{0r}},$$

which is asymptotically $\chi^2(1)$ under H_0.

When there is evidence of non-normality, T_n is no longer asymptotically $\chi^2(m)$ in general, but rather a scaled quadratic form in non-normal variables. For this case, we can correct the test statistic so that for large n, $E_{\theta_0}(T_n)$ behaves as $m\left(1 + \sigma_0^{-4}\kappa_4 \sum_{i=1}^n \left[V_0\left(V_0'V_0\right)^{-1}V_0'\right]_{i,i}^2\right)^{-1/2}$ and its variance equals $2m$, as the variance of a $\chi^2(m)$ variate. To do so, note that

$$Var_{\theta_0}\left(y'S_0'V_0\left(V_0'V_0\right)^{-1}V_0'S_0 y\right) = 2\sigma_0^2 m + \kappa_4 \sum_{i=1}^n \left[V_0\left(V_0'V_0\right)^{-1}V_0'\right]_{i,i}^2.$$

The non-normality corrected statistic is then given by

$$T_n' = \frac{T_n}{\left(1 + \frac{\hat{\kappa}_4}{2m\hat{\sigma}_0^4}\sum_{i=1}^n \left[\hat{V}_0\left(\hat{V}_0'\hat{V}_0\right)^{-1}\hat{V}_0'\right]_{i,i}^2\right)^{1/2}},$$

where $\hat{\kappa}_4$ is the sample fourth order cumulant of $\hat{S}_0 y$. Under normality, $\hat{\kappa}_4 \to_p 0$ and T_n' is asymptotically $\chi^2(m)$. Otherwise, the asymptotic behavior of the statistic depends on the sequence $\left\{\sum_{i=1}^n \left[\hat{V}_0\left(\hat{V}_0'\hat{V}_0\right)^{-1}\hat{V}_0'\right]_{i,i}^2\right\}$. In the scalar case, the corrected statistic collapses to

$$T_n' = \frac{\left(\hat{v}_{0r}'\hat{S}_0 y\right)^2}{\hat{\sigma}_0^2\left(\hat{v}_{0r}'\hat{v}_{0r}\right)\left(1 + \frac{\hat{\kappa}_4}{2\hat{\sigma}_0^4}\left(\hat{v}_{0r}'\hat{v}_{0r}\right)^{-1}\right)^{1/2}}.$$

In the next section we shall use the test in an empirical application.

6 An Empirical Application

The data is comprised of closing prices, in U.S. dollars, of 10 stocks which are dually-listed in the Tokyo Stock Exchange and in NASDAQ. Trading takes place in Tokyo between 7PM and 9PM and between 1030PM and 1AM, Eastern Standard Time (EST), whereas in New York trading runs 930AM through to 4PM. Therefore, there is no overlap in trading between the two markets. Daily data over 2003-2009 are available for most of the series. A precise summary of the data range and number of observations for each stock is given in Table 1. All data was transformed by a natural logarithm.

In our model-setting, Y_t and X_t are the logarithms of the closing prices in New York and in Tokyo, in U.S. dollars, respectively. On any given day, Y_t and X_t are recorded at 4PM and 1AM, EST, respectively, and when NASDAQ opens for trading, X_t is already known. Conceptually, the model for Y_t is constructed given the X_t-values, in exactly the same way which is done in the classical regression model. Augmented Dickey-Fuller (ADF) unit-root tests for Y_t, including all the common options of an intercept, trend and intercept and no trend and no intercept, are provided in Table 2. The p-values are notably large for all series in levels, whereas in first differences all p-values are zero. For this reason, the latter are not given in the table. Based on Table 2, it is tempting to simply fit the unit-root with a drift model (2) for the data.

Estimation results for model (11) and for model (1), with β_t given by (3) and (4), are given in Table 3. We denote the three models by M1, M2 and M3, respectively. For $\hat{\mu}$ of all models and for $\hat{w}_1$ of M1, the ordinary t-test that the parameter value equals zero is legitimate and so, the p-values, given in brackets, are based on it. On the other hand, the p-values associated with $\hat{w}_1$ of M2, of $\hat{w}_1$ and $\hat{w}_2$ of M3, and of the test of M2 vs. M3, are based on our asymptotic score-based T_n-test. Specifically, for a test of M1 vs. M2, which it will be convenient to denote by $T_{n1,2}$, we note that

$$[C]_{t,j} = \begin{cases} \exp\left(w_1 \Delta X_t\right), \text{ if } t-j=1 \\ 0, \text{ otherwise,} \end{cases}$$

so that

$$\left[\dot{S}_{02}\right]_{t,j} = \begin{cases} -\Delta X_t, \text{ if } t-j=1 \\ 0, \text{ otherwise.} \end{cases}.$$

In this case it holds that $\left[S_0^{-1}\right]_{t,j} = 1$, if $t \geq j$ and $\left[S_0^{-1}\right]_{t,j} = 0$, otherwise. For a simple test of the form $H_0 : w_j = 0$, $j = 1$ or $j = 2$, in M3, we have

$$\left[\dot{S}_{02}\right]_{t,j} = \begin{cases} -\Delta X_t, \text{ if } \Delta X_t > 0 \text{ and } t - j = 1 \\ 0, \text{ otherwise,} \end{cases}$$

or

$$\left[\dot{S}_{03}\right]_{t,j} = \begin{cases} -\Delta X_t, \text{ if } \Delta X_t \leq 0 \text{ and } t - j = 1 \\ 0, \text{ otherwise.} \end{cases}$$

Finally, for a test of M2 vs. M3, which we denote by $T_{n2,3}$, the required elements are

$$\left[\dot{S}_{02}\right]_{t,j} = \begin{cases} -\Delta X_t 1\left\{\Delta X_t > 0\right\} \exp(w_1 \Delta X_t), \text{ if } t - j = 1 \\ 0, \text{ otherwise.} \end{cases}$$

and

$$\left[\dot{S}_{03}\right]_{t,j} = \begin{cases} -\Delta X_t 1\left\{\Delta X_t \leq 0\right\} \exp(w_1 \Delta X_t), \text{ if } t - j = 1 \\ 0, \text{ otherwise.} \end{cases}$$

For M1, $\hat{w}_1$ is very close to unity throughout. For M2, $\hat{w}_1$ is positive with p-values equal to zero for all series and with a sample mean value of 0.21 across the 10 series. The positive sign of $\hat{w}_1$ implies that when $\Delta X_t > 0$, ΔY_t is expected to lie above the drift term. From a similarity perspective, Y_t adapts to a value which is consistent with its similarity to Y_{t-1}, through the closeness of X_t to X_{t-1}. For M3, all $\hat{w}_1$ and $\hat{w}_2$ values are positive but in general, $\hat{w}_1 > \hat{w}_2$. The meaning of this empirical result is that when $\Delta X_t > 0$, ΔY_t is expected to be greater than $\hat{\mu}$ and when $\Delta X_t \leq 0$, ΔY_t is expected to lie under $\hat{\mu}$, but the former effect is stronger than the latter. It should be noted though that the $T_{n2,3}$-test for the hypothesis $H_0 : w_1 = w_2$, could not be rejected for 4 out of the 10 cases.

We move on to discuss the goodness of fit of the different models, using mean squared error (MSE), the Akaike (AIC) and Schwarz (SC) criteria. For all series examined, the MSE measure is substantially lower for M2 and M3 compared with M1. The average reduction in the measure is 33% across the sample. The AIC and SC criteria are always smaller in M2 and M3 compared with M1 and between M2 and M3 the results are mixed, in line with the results for the $T_{n2,3}$-test.

Summary statistics for $\hat{\beta}_t$ from M2 and M3 are provided in Table 4. For all series, the means are close to unity, as expected, with ranges between 0.982 and 1.015. The explanation for the superior performance of M2 and M3 compared with M1 becomes particularly vivid when considering the variation of $\hat{\beta}_t$, as displayed in the case of Honda, in Figures 1 and 2. Other cases look very similar. Both graphs clearly fluctuate over the unit root, depending on the value of $\hat{\beta}_t$ for each t, but with a much larger volatility apparent over the latter period, prior to April 2009. It is this embedded variation over the whole sample-period which makes the difference in the goodness of fit.

We turn to discuss the results in Table 5, where we investigated further the performance of the score-based T_n-tests in our models. We denote by $T'_{n1,2}$ and $T'_{n2,3}$ the analogues of $T_{n1,2}$ and $T_{n2,3}$ with corrections for possible nonnormality. In addition, $T^s_{n1,2}$ is $T_{n1,2}$, calculated on simulated y-data and original x-data under M1, with parameter values given by the QMLE's of this model for each n. Similarly, $T^s_{n2,3}$ is $T_{n2,3}$, except the y-data was simulated under M2. Finally, $T'^s_{n1,2}$ and $T'^s_{n2,3}$ are $T^s_{n1,2}$ and $T^s_{n2,3}$ with corrections for possible nonnormality. Ordinary t-tests can be used for all other

The purpose of calculating $T^s_{n,1,2}$ and $T^s_{n2,3}$ is to show that the statistic delivers the right answer when the simulated data is consistent with M1 and M2, respectively. The data set considered for this table is comprised of the first 250, 500 and 1000 observations of each series, apart from Mizuho for which a total of only 579 observations is available. Apart for Mizuho then, this data set excludes the more volatile period of 2008-2009 from the analysis. Clearly, for all data and sample sizes considered, $T_{n1,2}$ is highly significant, whereas the opposite holds for $T^s_{n,1,2}$, as expected. Hence, M1 is rejected in favor of M2 emphatically in the real data, whereas the converse holds for the simulated data, as expected. For the test of M2 vs. M3, both $T_{n2,3}$ and $T^s_{n2,3}$ are usually insignificant and so, we cannot reject M2 in favor of M3. Finally, the tests with nonnormality correction are not substantially different from the original tests, which means that the correction is generally not essential for this data set. Overall, the new score-based T_n-test delivers the answers expected and in general agreement with the other results given in Table 3.

7 Remarks

There is already a great deal of literature on what has been termed the "embarrassing resilience of unit root tests", see for instance, the comprehensive discussion by Hayashi (2000) and the references therein. Figures 1 and 2 provide what might be a typical view of a data set for which the unit root hypothesis cannot be rejected, but clearly the unit root holds only *on average.* Around this average, the coefficient can fluctuate between stationary- and explosive-type behavior, depending on the similarity between Y_t and its past. It appears that this embedded variation can lead to substantial improvements in goodness of fit. A strong feature of the model, is that the idea behind it is axiomatically justified, see Gilboa *et. al.* (2006).

The paper can be extended in a number of directions, notably, higher order autoregressions and dependent error processes, but we believe that the main idea is sufficiently transparent in the model considered here. While, the asymptotic theory of estimation and testing is nonstandard, due to the fact that the model is in general nonstationary, the practical implementation of the procedures can be carried out on any standard econometric software. Overall, , as we have seen in the empirical section, this approach has much to commend in it.

References

Cai, Z. (2002) Two-step likelihood estimation procedure for varying-coefficient models. *Journal of Multivariate Analysis* 82, 189–209.

Cai, Z. (2007) Trending time varying coefficient time series models with serially correlated errors. *Journal of Econometrics* 136, 163-188.

Cai, Z., J. Fan, & R. Li (2000) Efficient estimation and inferences for varying-coefficient models, *Journal of the American Statistical Association* 95, 888–902.

Cai, Z., & Q. Li (2008) Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory* 24, 1321-1342.

Cai, Z., Q. Li, & J. Park, (2009) Functional coefficient models for nonstationary time series data. *Journal of Econometrics* 148, 101-113.

Chen, R. & R. S. Tsay (1993) Functional-coefficient autoregressive models. *Journal of the American Statistical Association* 88, 298–308.

Dahlhaus, R., R.H. Neumann & R. Von Sachs (1999) Nonlinear wavelet estimation of time-varying autoregressive process. *Bernoulli* 5, 873–906.

Evans, G.B.A. & N.E. Savin (1984) Testing for Unit Roots: 2. *Econometrica* 52, 1241–1269.

Fan, J. & W. Zhang (1999) Statistical estimation in varying-coefficient models. *The Annals of Statistics* 27, 1491–1518.

Gayer, G., I. Gilboa & O. Lieberman (2007) Rule-based and case-based reasoning in real estate prices. *The B.E. Journals in Theoretical Economics* 7, No. 1 (Advances), Article 10.

Gilboa, I., O. Lieberman & D. Schmeidler (2006) Empirical similarity. *The Review of Economics and Statistics* 88, 433–444.

Gilboa, I., O. Lieberman & D. Schmeidler (2010) A similarity-based approach to prediction. *Journal of Econometrics*. Forthcoming.

Graybill, F.A. (1983) *Matrices with Applications in Statistics*, 2nd Edition. Wadsworth, Belmont, California.

Hastie, T.J. & R.J. Tibshirani (1993) Varying-coefficient models (with discussion). J*Journal of the Royal Statistical Society* Series B 55, 757–796.

Hayashi, F. (2000) *Econometrics*. Princeton University Press.

Hoderlein, S., J. Klemelä & E. Mammen (2010) Analyzing the random coefficient model nonparametrically. *Econometric Theory* 26, 804-837.

Lieberman, O. (1997) The effect of nonnormality. *Econometric Theory* 13, 52–78.

Lieberman, O. (2010) Asymptotic Theory for Empirical Similarity models. *Econometric Theory*. Forthcoming.

Lundbergh, S., T. Terǎsvirta, & D. van Dijk (2003) Time-varying smooth transition autoregressive models. *Journal of Business and Economic Statistics* 21, 104-121.

Nicholls, D.F. & B.G. Quinn (1980) The estimation of random coefficient autoregressive models. I. *Journal of Time Series Analysis* 1, 37– 46.

Nicholls, D.F. and B.G. Quinn (1981) The estimation of random coefficient autoregressive models. II. *Journal of Time Series Analysis* 2, 185–203.

Nicholls, D.F. & B.G. Quinn (1982) *Random Coefficient Autoregressive Models: An Introduction*. Lecture Notes in Statistics 11. Springer-Verlag.

Park, J., & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69, 117-162.

Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.

Wang, Q. Y., & P.C.B. Phillips (2009a) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710-738.

Wang, Q. Y., & P.C.B. Phillips (2009b) Structural nonparametric cointegrating regression. *Econometrica* 77, 1901-1948.

Wu, C.F. (1981) Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics* 9, 501–513.

Xiao, Z. (2009) Functional coefficient cointegration models. *Journal of Econometrics* 152, 81-92.

Table1. Data description for the dual-stock data

	Honda	**Canon**	**NTT DoCoMo**	**Fujifilm**	**Toyota**
Starting Date	12-Aug-02	6-Jan-03	6-Jan-03	6-Jan-03	6-Jan-03
Ending date	19-Mar-09	24-Apr-09	24-Apr-09	24-Apr-09	24-Apr-09
Observations	1564	1496	1496	1496	1496
	Mizuho	**Mitsubishi**	**Kyocera**	**Nissan**	**Nomura**
Starting Date	8-Nov-06	17-May-02	6-Jan-03	20-Feb-03	6-Jan-03
Ending date	21-Apr-09	19-Mar-09	24-Apr-09	19-Mar-09	24-Apr-09
Observations	579	1622	1496	1437	1496

Table 2. p-values of the ADF tests for the dual-stock data

	Honda	**Canon**	**NTT DoCoMo**	**FujiFilm**	**Toyota**
Intercept	0.3061	0.3317	0.1845	0.4047	0.4191
Trend and intercept	0.5491	0.8547	0.0456	0.7473	0.8716
None	0.6788	0.7030	0.4755	0.5613	0.7969
	Mizuho	**Mitsubishi**	**Kyocera**	**Nissan**	**Nomura**
Intercept	0.8926	0.6168	0.1048	0.9693	0.8658
Trend and intercept	0.2222	0.9433	0.2594	0.9482	0.9586
None	0.1278	0.4958	0.7363	0.3618	0.4395

Note: Intercept-the ADF test with an intercept; Trend and intercept-the ADF test with trend and intercept; None-the ADF test without a trend or an intercept.

Table 3. Similarity-based model estimation of the dual-stock data

		$\hat{\mu}$	$\hat{w}_1$	$\hat{w}_2$	**RSS/n**	**AIC**	**SC**	$T_{n2,3}$
	M1	0.0153	0.9953		0.0005	-4.8032	-4.7963	
		(0.0500)	(0.0000)					
Honda	M2	0.0000	0.1851		0.0003	-5.3818	-5.3750	
		(0.9826)	(0.0000)					
	M3	-0.0004	0.1928	0.1768	0.0003	-5.3813	-5.3710	3.9438
		(0.4585)	(0.0000)	(0.0000)				(0.1392)
	M1	0.0155	0.9958		0.0005	-4.8446	-4.8375	
		(0.0563)	(0.0000)					
Canon	M2	0.0000	0.1624		0.0003	-5.3245	-5.3174	
		(0.9249)	(0.0000)					
	M3	0.0000	0.1631	0.1618	0.0003	-5.3232	-5.3125	1.3051
		(0.9977)	(0.0000)	(0.0000)				(0.5207)
	M1	0.0251	0.9911		0.0004	-4.9194	-4.9123	
		(0.0117)	(0.0000)					
NTT DoCoMo	M2	-0.0001	0.2412		0.0002	-5.6070	-5.5999	
		(0.8347)	(0.0000)					
	M3	-0.0005	0.2502	0.2332	0.0002	-5.6062	-5.5955	1.6558
		(0.4164)	(0.0000)	(0.0000)				(0.4370)
	M1	0.0216	0.9938		0.0005	-4.8598	-4.8527	
		(0.0393)	(0.0000)					
FujiFilm	M2	-0.0001	0.1957		0.0002	-5.4847	-5.4776	
		(0.8858)	(0.0000)					
	M3	-0.0016	0.2249	0.1678	0.0002	-5.4919	-5.4812	19.4878
		(0.0064)	(0.0000)	(0.0000)				(0.0001)
	M1	0.0137	0.9970		0.0004	-5.0234	-5.0163	
		(0.0798)	(0.0000)					
Toyota	M2	0.0001	0.1229		0.0003	-5.4016	-5.3945	
		(0.7691)	(0.0000)					
	M3	-0.0011	0.1418	0.1024	0.0003	-5.4060	-5.3954	12.2765
		(0.0573)	(0.0000)	(0.0000)				(0.0022)

Note: M1-the random walk with a drift model; For M1, $\hat{w}_1 \equiv \hat{\beta}$; M2-the similarity model with β_t as in (3); M3-the similarity model with β_t as in (4); $T_{n2,3}$ - the T_n-test for $w_1 = w_2$ in M3; p-values in brackets.

Table 3 (continued). Similarity-based model estimation of the dual-stock data

		$\hat{\mu}$	$\hat{w}_1$	$\hat{w}_2$	**RSS/n**	**AIC**	**SC**	$T_{n2,3}$
	M1	0.0009	0.9985		0.0015	-3.6646	-3.6495	
		(0.9186)	(0.0000)					
Mizuho	M2	-0.0012	0.2506		0.0012	-3.9160	-3.9009	
		(0.3995)	(0.0000)					
	M3	-0.0036	0.2951	0.2072	0.0012	-3.9165	-3.8939	3.6799
		(0.0915)	(0.0000)	(0.0000)				(0.1588)
	M1	0.0056	0.9973		0.0008	-4.2876	-4.2809	
		(0.2056)	(0.0000)					
Mitsubishi	M2	-0.0001	0.3311		0.0005	-4.8651	-4.8584	
		(0.8039)	(0.0000)					
	M3	-0.0008	0.3460	0.3140	0.0005	-4.8647	-4.8547	14.4319
		(0.3069)	(0.0000)	(0.0000)				(0.0007)
	M1	0.0347	0.9920		0.0005	-4.8403	-4.8332	
		(0.0107)	(0.0000)					
Kyocera	M2	0.0000	0.1455		0.0003	-5.3928	-5.3857	
		(0.9059)	(0.0000)					
	M3	-0.0016	0.1685	0.1224	0.0003	-5.4005	-5.3899	14.7001
		(0.0101)	(0.0000)	(0.0000)				(0.0006)
	M1	-0.0013	1.0003		0.0005	-4.8509	-4.8439	
		(0.8074)	(0.0000)					
Nissan	M2	-0.0005	0.2070		0.0003	-5.3008	-5.2931	
		(0.3071)	(0.0000)					
	M3	-0.0005	0.1944	0.2085	0.0002	-5.5648	-5.5527	15.1118
		(0.2653)	(0.0000)	(0.0000)				(0.0005)
	M1	0.0049	0.9980		0.0008	-4.3148	-4.3077	
		(0.4503)	(0.0000)					
Nomura	M2	-0.0002	0.2596		0.0005	-4.8230	-4.8159	
		(0.7179)	(0.0000)					
	M3	-0.0021	0.2952	0.2226	0.0005	-4.8274	-4.8167	10.6317
		(0.0149)	(0.0000)	(0.0000)				(0.0049)

Note: M1-the random walk with a drift model; For M1, $\hat{w}_1 \equiv \hat{\beta}$; M2-the similarity model with β_t as in (3); M3-the similarity model with β_t as in (4); $T_{n2,3}$ - the T_n-test for $w_1 = w_2$ in M3; p-values in brackets.

Table 4. Summary statistics for $\hat{\beta}_t$

		Honda	Canon	NTT DoCoMo	FujiFilm	Toyota
M2	Mean	1.0000	1.0000	1.0000	1.0000	1.0000
	Standard Deviation	0.0020	0.0016	0.0022	0.0019	0.0011
	Minimum	0.9882	0.9912	0.9834	0.9891	0.9942
	Maximum	1.0120	1.0096	1.0088	1.0114	1.0063
M3	Mean	1.0001	1.0000	1.0000	1.0002	1.0001
	Standard Deviation	0.0020	0.0016	0.0022	0.0019	0.0011
	Minimum	0.9887	0.9912	0.9840	0.9906	0.9952
	Maximum	1.0125	1.0097	1.0091	1.0131	1.0072
		Mizuho	**Mitsubishi**	**Kyocera**	**Nissan**	**Nomura**
M2	Mean	0.9998	1.0000	1.0000	1.0000	1.0000
	Standard Deviation	0.0040	0.0030	0.0014	0.0021	0.0030
	Minimum	0.9820	0.9822	0.9907	0.9881	0.9820
	Maximum	1.0185	1.0157	1.0119	1.0127	1.0146
M3	Mean	1.0003	1.0004	1.0002	0.9999	1.0003
	Standard Deviation	0.0041	0.0031	0.0015	0.0021	0.0030
	Minimum	0.9851	0.9852	0.9922	0.9880	0.9845
	Maximum	1.0218	1.0185	1.0138	1.0120	1.0166

Note: M2-the similarity model with $d_w(x_t, x_{t-1})$ as in (3); M3-the similarity model with $d_w(x_t, x_{t-1})$ as in (4).

Table 5. Score-based tests for M1 vs. M2 and for M2 vs. M3

	n	$T_{n1,2}$	$T'_{n1,2}$	$T^s_{n1,2}$	$T'^s_{n1,2}$	$T_{n2,3}$	$T'_{n2,3}$	$T^s_{n2,3}$	$T'^s_{n2,3}$
Honda	250	103.149	102.405	0.2425	0.2428	0.4571	0.4494	1.4386	1.4390
		(0.0000)	(0.0000)	(0.8858)	(0.8857)	(0.7957)	(0.7988)	(0.4871)	(0.4870)
	500	206.046	205.605	1.0204	1.0213	0.2346	0.2335	0.8166	0.8156
		(0.0000)	(0.0000)	(0.6004)	(0.6001)	(0.8893)	(0.8898)	(0.6648)	(0.6651)
	1000	360.464	359.558	0.4228	0.4230	5.0504	5.0162	1.9154	1.9162
		(0.0000)	(0.0000)	(0.8095)	(0.8094)	(0.0800)	(0.0814)	(0.3838)	(0.3836)
Canon	250	113.043	113.16	0.4382	0.4381	0.3600	0.3611	0.3801	0.3805
		(0.0000)	(0.0000)	0.8032	(0.8033)	(0.8353)	(0.8348)	(0.8269)	(0.8268)
	500	227.799	227.845	2.3422	2.3411	1.9456	1.9463	0.7049	0.7041
		(0.0000)	(0.0000)	(0.3100)	(0.3102)	(0.3780)	(0.3779)	(0.7030)	(0.7032)
	1000	388.6	388.464	0.0181	0.0181	0.6887	0.6874	1.2390	1.2375
		(0.0000)	(0.0000)	(0.9910)	(0.9910)	(0.7087)	(0.7091)	(0.5382)	(0.5286)
NTT-Docomo	250	139.198	138.397	2.2769	2.2772	1.6810	1.6732	0.4653	0.4652
		(0.0000)	(0.0000)	(0.3203)	(0.3203)	(0.4315)	(0.4332)	(0.7924)	(0.7924)
	500	263.401	260.759	2.8078	2.8086	4.3416	4.1709	0.8113	0.8067
		(0.0000)	(0.0000)	(0.2456)	(0.2455)	(0.1141)	(0.1243)	(0.6655)	(0.6681)
	1000	414.082	412.743	0.8106	0.8107	1.2521	1.2384	1.1548	1.1553
		(0.0000)	(0.0000)	(0.6668)	(0.6667)	(0.5347)	(0.5384)	(0.5614)	(0.5612)
Fuji-Film	250	120.669	120.393	0.0019	0.0019	1.6067	1.6116	0.4950	0.4957
		(0.0000)	(0.0000)	(0.9991)	(0.9991)	(0.4478)	(0.4467)	(0.7808)	(0.7805)
	500	226.014	225.649	0.0163	0.0163	0.6343	0.6332	5.4087	5.4055
		(0.0000)	(0.0000)	(0.9919)	(0.9919)	(0.7282)	(0.7286)	(0.0669)	(0.0670)
	1000	396.656	396.176	2.6995	2.6997	3.2050	3.1990	0.1511	0.1513
		(0.0000)	(0.0000)	(0.2632)	(0.2593)	(0.2014)	(0.2020)	(0.9272)	(0.9271)
Toyota	250	116.364	115.405	4.9561	4.9728	4.9656	4.8922	1.7381	1.7478
		(0.0000)	(0.0000)	(0.0839)	(0.0832)	(0.0835)	(0.0866)	(0.4193)	(0.4173)
	500	225.262	224.684	5.3971	5.4018	0.7877	0.7843	1.0873	1.0869
		(0.0000)	(0.0000)	(0.0673)	(0.0671)	(0.6745)	(0.6756)	(0.5832)	(0.5807)
	1000	360.307	359.526	0.4285	0.4286	3.5118	3.4921	0.2937	0.2938
		(0.0000)	(0.0000)	(0.8071)	(0.8071)	(0.1728)	(0.1745)	(0.8634)	(0.8634)

Note: $T_{n1,2}$ and $T_{n2,3}$ are score-based tests for M1 vs. M2 and for M2 vs. M3, respectively. $T'_{n1,2}$ and $T'_{n2,3}$ are $T_{n1,2}$ and $T_{n2,3}$ with corrections for possible nonnormality. $T^s_{n1,2}$ is $T_{n1,2}$ calculated on simulated y-data under M1 with parameter values given by the QMLE's of this model for each n. Similarly for $T^s_{n2,3}$, where the y-data was simulated under M2. Finally, $T'^s_{n1,2}$ and $T'^s_{n2,3}$ are $T^s_{n1,2}$ and $T^s_{n2,3}$ with corrections for possible nonnormality.

Table 5 (continued). Score-based tests for M1 vs. M2 and for M2 vs. M3

	n	$T_{n1,2}$	$T'_{n1,2}$	$T^s_{n1,2}$	$T'^s_{n1,2}$	$T_{n2,3}$	$T'_{n2,3}$	$T^s_{n2,3}$	$T'^s_{n2,3}$
	250	114.995	113.623	1.7054	1.7083	0.9357	0.9170	7.8458	7.8442
		(0.0000)	(0.0000)	(0.4263)	(0.4256)	(0.6263)	(0.6322)	(0.0198)	(0.0198)
Mizuho	500	179.582	175.692	3.8998	3.9027	6.8275	6.5197	5.2222	5.1758
		(0.0000)	(0.0000)	(0.1423)	(0.1421)	(0.0329)	(0.0384)	(0.0735)	(0.0752)
	250	131.583	127.389	0.0132	0.0132	7.5090	6.8515	0.8046	0.8054
		(0.0000)	(0.0000)	(0.9934)	(0.9934)	(0.0234)	(0.0325)	(0.6688)	(0.6685)
Mitsubishi	500	260.662	258.541	0.0822	0.0822	0.6444	0.6285	1.5801	1.5810
		(0.0000)	(0.0000)	(0.9597)	(0.9597)	(0.7246)	(0.7303)	(0.4538)	(0.4536)
	1000	393.284	391.134	0.9885	0.9889	13.4667	13.2743	0.9252	0.9256
		(0.0000)	(0.0000)	(0.6100)	(0.6099)	(0.0012)	(0.0013)	(0.6296)	(0.6295)
	250	88.409	87.767	1.1435	1.1345	3.2607	3.2271	2.6415	2.6363
		(0.0000)	(0.0000)	(0.5645)	(0.5671)	(0.1959)	(0.1992)	(0.2669)	(0.2676)
Kyocera	500	201.201	200.659	0.2947	0.2943	0.1418	0.1413	1.6070	1.6062
		(0.0000)	(0.0000)	(0.8630)	(0.8632)	(0.9316)	(0.9318)	(0.4478)	(0.4479)
	1000	379.973	378.529	2.1421	2.1413	0.9786	0.9713	7.2443	7.2354
		(0.0000)	(0.0000)	(0.3426)	(0.3428)	(0.6131)	(0.6153)	(0.0267)	(0.0268)
	250	91.661	91.429	0.0518	0.0519	1.3797	1.3737	0.8485	0.8506
		(0.0000)	(0.0000)	(0.9744)	(0.9744)	(0.5017)	(0.5032)	(0.6543)	(0.6536)
Nissan	500	179.458	179.17	1.7347	1.7348	0.8989	0.8966	1.1008	1.0999
		(0.0000)	(0.0000)	(0.4201)	(0.4200)	(0.6380)	(0.6387)	(0.5767)	(0.5770)
	1000	280.993	278.95	1.7253	1.7266	6.0082	5.8378	0.7241	0.7259
		(0.0000)	(0.0000)	(0.4220)	(0.4218)	(0.7405)	(0.0540)	(0.6962)	(0.6956)
	250	125.286	124.772	0.0700	0.0701	0.6548	0.6553	2.5945	2.5894
		(0.0000)	(0.0000)	(0.9656)	(0.9656)	(0.7208)	(0.7206)	(0.2733)	(0.2740)
Nomura	500	218.313	217.853	0.5567	0.5569	0.4929	0.4919	0.3155	0.3157
		(0.0000)	(0.0000)	(0.7570)	(0.7570)	(0.7816)	(0.7820)	(0.8541)	(0.8540)
	1000	429.08	428.476	0.8092	0.8092	3.2508	3.2332	1.8709	1.8718
		(0.0000)	(0.0000)	(0.6672)	(0.6672)	(0.1968)	(0.1986)	(0.3924)	(0.3922)

Note: $T_{n1,2}$ and $T_{n2,3}$ are score-based tests for M1 vs. M2 and for M2 vs. M3, respectively. $T'_{n1,2}$ and $T'_{n2,3}$ are $T_{n1,2}$ and $T_{n2,3}$ with corrections for possible nonnormality. $T^s_{n1,2}$ is $T_{n1,2}$ calculated on simulated y-data under M1 with parameter values given by the QMLE's of this model for each n. Similarly for $T^s_{n2,3}$, where the y-data was simulated under M2. Finally, $T'^s_{n1,2}$ and $T'^s_{n2,3}$ are $T^s_{n1,2}$ and $T^s_{n2,3}$ with corrections for possible nonnormality.

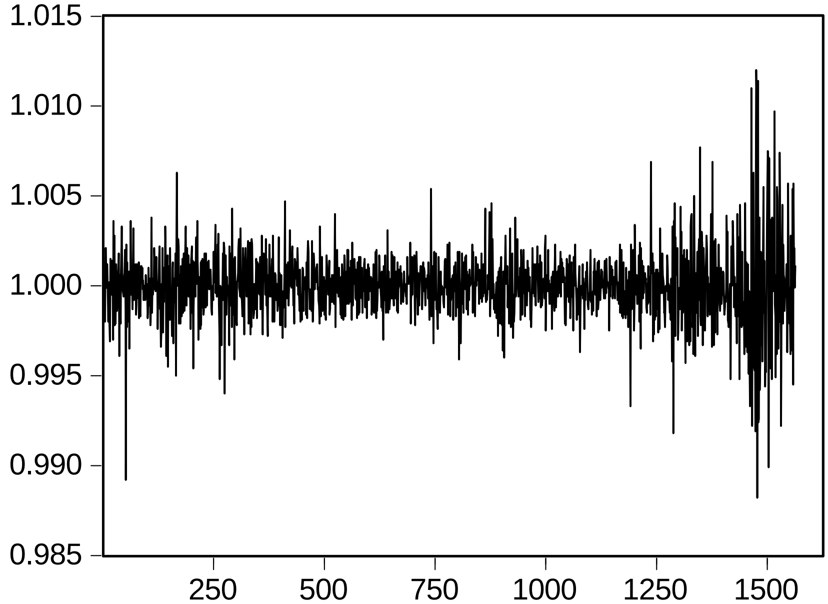

Figure 1: $\hat{\beta}_t$ from M2 in the Honda series

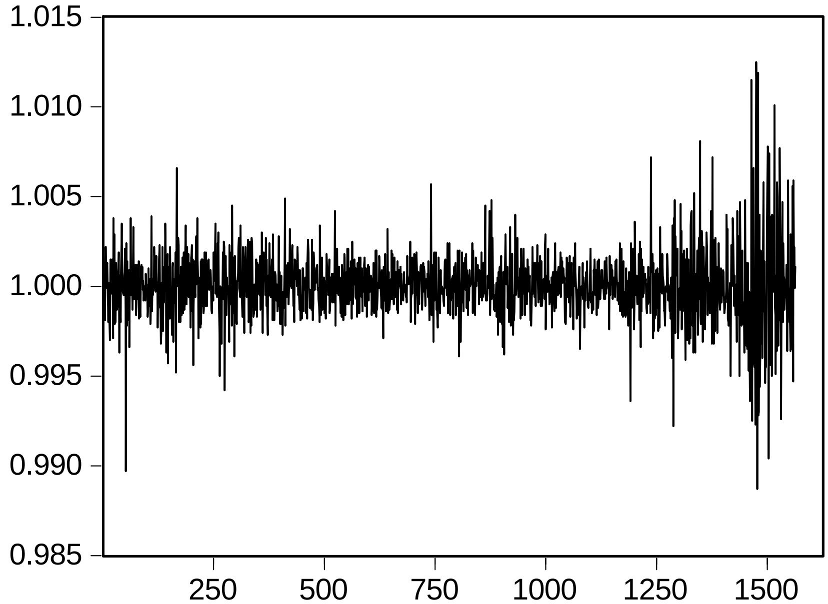

Figure 2: $\hat{\beta}_t$ from M3 in the Honda series