BINGHAM McCUTCHEN LLP
DONN P. PICKETT (SBN 72257)
GEOFFREY M. HOWARD (SBN 157468)
HOLLY A. HOUSE (SBN 136045)
ZACHARY J. ALINDER (SBN 209009)
BREE HANN (SBN 215695)
Three Embarcadero Center
San Francisco, CA 94111-4067
Telephone: (415) 393-2000
Facsimile: (415) 393-2286
donn.pickett@bingham.com
geoff.howard@bingham.com
holly.house@bingham.com
zachary.alinder@bingham.com
bree.hann@bingham.com

BOIES, SCHILLER & FLEXNER LLP
DAVID BOIES (Admitted *Pro Hac Vice*)
333 Main Street
Armonk, NY 10504
Telephone: (914) 749-8200
Facsimile: (914) 749-8300
dboies@bsfllp.com
STEVEN C. HOLTZMAN (SBN 144177)
FRED NORTON (SBN 224725)
1999 Harrison St., Suite 900
Oakland, CA 94612
Telephone: (510) 874-1000
Facsimile: (510) 874-1460
sholtzman@bsfllp.com
fnorton@bsfllp.com

DORIAN DALEY (SBN 129049)
JENNIFER GLOSS (SBN 154227)
500 Oracle Parkway, M/S 5op7
Redwood City, CA 94070
Telephone: (650) 506-4846
Facsimile: (650) 506-7114
dorian.daley@oracle.com
jennifer.gloss@oracle.com

Attorneys for Plaintiffs Oracle USA, Inc., *et al.*

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
OAKLAND DIVISION

ORACLE USA, INC., *et al.*,

Plaintiffs,

v.

SAP AG, *et al.*,

Defendants.

CASE NO. 07-CV-01658 PJH (EDL)

EXHIBIT 6 TO THE DECLARATION OF DANIEL S. LEVY, PH.D. IN SUPPORT OF MOTION NO. 1: TO EXCLUDE TESTIMONY OF DEFENDANTS' EXPERT STEPHEN CLARKE

FILED PURSUANT TO DKT. NO. 915

EXHIBIT 6

Estimation and Inference in Econometrics

RUSSELL DAVIDSON
JAMES G. MACKINNON

New York Oxford
OXFORD UNIVERSITY PRESS
1993

Offering
tive not r
this inno
rical arg
intuitive
advanced
sizing th
tions of
technique
the text is
for estim
tion testi
ing diag
constancy
dasticity,
tion. Exp
obtained
authors
descriptio
translate
standard e
Coverin
problems
those that
accessible
vital topi
indispensa
econometr
interested
Flexibly
variety of
complete
and separa
ized interes

Oxford University Press

Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Kuala Lumpur Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland Madrid

and associated companies in
Berlin Ibadan

Published by Oxford University Press, Inc.,
198 Madison Avenue, New York, New York 10016-4314

Oxford is a registered trademark of Oxford University Press

Library of Congress Cataloging-in-Publication Data
Davidson, Russell.
Estimation and inference in econometrics /
by Russell Davidson and James G. MacKinnon.
p. cm.
Includes bibliographical references and index.
ISBN 0-19-506011-3
1. Econometrics. I. MacKinnon, James G.
II. Title. HB139.D368 1993
330'.01'5195—dc20 92-12048

6 8 9 7 5

Printed in the United States of America
on acid-free paper

trend-stationary, that is, stationary around a trend. In contrast, the second model, (19.02), says that y_t follows a **random walk with drift**. The drift parameter δ_1 in (19.02) plays much the same role as the trend parameter γ_2 in (19.01), since both cause y_t to trend upward over time. But the behavior of y_t is very different in the two cases, because in the first case detrending it will produce a variable that is stationary, while in the second case it will not.

There has been a great deal of literature on which of these two models, the trend-stationary model (19.01) or the random walk with drift (19.02), best characterizes most economic time series. Nelson and Plosser (1982) is a classic paper, Campbell and Mankiw (1987) is a more recent one, and Stock and Watson (1988a) provides an excellent discussion of many of the issues. In the next chapter we will discuss some of the methods that can be used to decide whether a given time series is well characterized by either of these models. For now, what concerns us is what happens if we use time series that are described by either of these two models as dependent or independent variables in a regression model.

If a time series with typical element x_t trends upward forever, then $n^{-1}\sum_{t=1}^{n} x_t^2$ will diverge to $+\infty$. Thus, if such a series is used as a regressor in a linear regression model, the matrix $n^{-1}\boldsymbol{X}^{\top}\boldsymbol{X}$ cannot possibly tend to a finite, positive definite matrix. All of the asymptotic theory we have used in this book is therefore inapplicable to models in which any of the regressors is well characterized by (19.01) or (19.02).[1] This does not mean that one should *never* put a trending variable on the right-hand side of a linear or nonlinear regression. Since the samples we actually observe are finite, and often quite small, we can never be sure that a series will trend upward forever. Moreover, the desirable finite-sample properties of least squares regression hold whether or not the regressors trend upward. But if we wish to rely on conventional asymptotic theory, it would seem to be prudent to specify our models so that strongly trending variables do not appear on the right-hand side. This in turn means that the dependent variable cannot be strongly trending. The most common approach is to take first differences of all such variables before specifying the model.

One compelling reason for taking first differences of trending variables is the phenomenon of **spurious regression**. It should be obvious that if two variables, say y_t and x_t, both trend upward, a regression of y_t on x_t is very

[1] The fact that *standard* asymptotic theory is inapplicable to such models does not mean that no such theory applies to them. For example, we studied a simple model of regression on a linear trend in Section 4.4 and found that the least squares estimator of the coefficient on the trend term was consistent, but with a variance that was $O(n^{-3})$ instead of the more conventional $O(n^{-1})$. Moreover, since there exist CLTs that apply to such models, the usual procedures for inference are asymptotically valid. For example, if $u_t \sim \text{IID}(0,\sigma^2)$ and $S_n \equiv n^{-3/2}\sum_{t=1}^{n} tu_t$, then S_n tends in distribution to $N(0,\sigma^2/3)$. Notice that the normalizing factor here is $n^{-3/2}$ rather than $n^{-1/2}$.

likely to find a "significant"
they have in common is the
of y_t on x_t and a constant w
can be characterized by (19.
the stochastic parts of y_t ar
this result and are advised t

It is intuitively very pl
icant, but actually spurious
both trend upward over tim
appears at first to be a mu
They considered time series
drift, that is, series generat
found, by Monte Carlo expe
dom walks, the t statistic fo

rejects the null hypothesis f
it more and more frequentl
Phillips (1986) proved that
the time, asymptotically.

Some Monte Carlo resul
Each column shows the prop
the t statistic for $\beta = 0$ in sc
5% level. For column 1, tl
generated by independent r
x_t and y_t are the same as f
added to the regression. For
again. For column 3, both
walks with drift, the drift pa
error σ (this ratio is the on
t statistic). For column 4, l
series, with the trend coeffic

The results in columns
x_t and y_t are both trending
results is how rapidly the nu
This is a consequence of th
information in the sample is
increasing faster in the tren
drift.

In contrast, the results i
After all, x_t and y_t are totall
So why do we often very c
of a relationship when we re

likely to find a "significant" relationship between them, even if the only thing they have in common is the upward trend. In fact, the R^2 for a regression of y_t on x_t and a constant will tend to unity as $n \rightarrow \infty$ whenever both series can be characterized by (19.01), even if there is no correlation at all between the stochastic parts of y_t and x_t. Readers may find it illuminating to prove this result and are advised to look at Section 4.4 for some useful results.

It is intuitively very plausible that we should observe apparently significant, but actually spurious, relationships between unrelated variables that both trend upward over time. Granger and Newbold (1974) discovered what appears at first to be a much more surprising form of spurious regression. They considered time series which are generated by **random walks without drift**, that is, series generated by a process like $y_t = y_{t-1} + u_t$. What they found, by Monte Carlo experiment, is that if x_t and y_t are independent random walks, the t statistic for $\beta = 0$ in the regression

$$y_t = \alpha + \beta x_t + u_t \tag{19.03}$$

rejects the null hypothesis far more often than it should and tends to reject it more and more frequently the larger is the sample size n. Subsequently, Phillips (1986) proved that this t statistic will reject the null hypothesis all the time, asymptotically.

Some Monte Carlo results on spurious regressions are shown in Table 19.1. Each column shows the proportion of the time, out of 10,000 replications, that the t statistic for $\beta = 0$ in some regression rejected the null hypothesis at the 5% level. For column 1, the regression is (19.03), and both x_t and y_t are generated by independent random walks with n.i.d. errors. For column 2, x_t and y_t are the same as for column 1, but a lagged dependent variable is added to the regression. For columns 3 and 4, the regression is simply (19.03) again. For column 3, both x_t and y_t are generated by independent random walks with drift, the drift parameter δ_1 being one-fifth the size of the standard error σ (this ratio is the only parameter that affects the distribution of the t statistic). For column 4, both x_t and y_t are independent trend-stationary series, with the trend coefficient γ_2 being 1/25 the size of σ.

The results in columns 3 and 4 of the table are not very surprising, since x_t and y_t are both trending upward. The only interesting thing about these results is how rapidly the number of rejections increases with the sample size. This is a consequence of the fact that, in both these cases, the amount of information in the sample is increasing at a rate faster than n. It is evidently increasing faster in the trend case than in the case of the random walk with drift.

In contrast, the results in columns 1 and 2 of the table may be surprising. After all, x_t and y_t are totally independent series, and neither contains a trend. So why do we often — very often indeed for large sample sizes — find evidence of a relationship when we regress y_t on x_t? One answer should be obvious to