Université de Montréal

Three Essays on Econometrics of Latent Variables

par

Joanna Jasiak

Département des sciences économiques

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiae Doctor (Ph.D.)
en sciences économiques

Octobre 1995

Page d'identification du jury

Université de Montréal

Faculté des études supérieures

Cette Thèse intitulée:

Three Essays on Econometrics of Latent Variables

présentée par

Joanna Jasiak

a été évaluée par un jury composé des personnes suivantes:

Thèse acceptée le: Lundi 6 mai 1996

André MARTENS .... président du jury

Eric GHYSELS .... directeur de recherche

Jean-Marie DUFOUR .... codirecteur de recherche

René GARCIA .... membre du jury

Eric RENAULT .... examinateur externe

Christiane ROUSSEAU .... représentante ou représentant du doyen

Summary

Data available to economists is often incomplete. Frequently values of the explanatory variables cannot be observed or are observed with a noise. To deal with the data limitation problems, various latent variable models have been considered in the econometric and statistic literature. In this thesis, we examine two categories of these models and propose new approaches to estimation and inference. The first category consists of regression models with unobserved explanatory variables, such as expectational variables. The second category contains models of financial time series with a latent process of stochastic volatility. In the first essay, we focus on inference in regression models while the stochastic volatility models are discussed in essays two and three.

Finite sample tests and confidence sets for models with unobserved and generated regressors as well as various models estimated by instrumental variable method are first proposed. We study two distinct approaches for various models considered by Pagan (1984). The first one is an instrument substitution method which generalizes an approach proposed by Anderson and Rubins (1949) and Fuller (1987) for different (although related) problems, while the second one is based on splitting the sample. The instrument substitution method uses the instruments directly, instead of generated regressors, in order to test hypotheses about the "structural parameters" of interest and build confidence sets. The second approach relies on "generated regressors", which allows a gain in degrees of freedom, and a sample split technique. A distributional theory is obtained under the assumptions of Gaussian errors and strictly exogenous regressors. We show that the various tests and confidence sets proposed are (locally) "asymptotically valid" under much weaker assumptions. The properties of the tests proposed are examined in simulation experiments. In general, they outperform the usual asymptotic inference methods in terms of both reliability and power. Finally, the techniques suggested are applied to a model of Tobin's q and to a model of academic performance.

In the second essay, we study stochastic volatility models with time deformation.

Such processes relate to early works by Mandelbrot and Taylor (1967), Clark (1975), Tauchen and Pitts (1983), among others. In our setup, the latent process of stochastic volatility evolves in an operational time which differs from calendar time. The time deformation can be determined by part volume of trade, part price changes, possibly with an asymmetric leverage effect, and other variables setting the pace of information arrival.

The econometric specification exploits the state-space approach for stochastic volatility models proposed by Harvey, Ruiz and Shephard (1994) as well as matching moment estimation procedures using SNP densities of stock returns and trading volume estimated by Gallant, Rossi and Tauchen (1992). Daily data on the price changes and volume of trade of the S&P500 over a 1950–1987 sample are investigated. Supporting evidence for a time deformation representation is found and its impact on the behavior of price series and volume is analyzed. We find that increases in volume accelerate operational time, resulting in volatility being less persistent and subject to shocks with a higher innovation variance. Downward price movements have similar effects while upward price movements increase persistence in volatility and decrease the dispersion of shocks by slowing down the operational time clock. We present the basic model as well as several extensions. In particular, we formulate and estimate a bivariate return-volume stochastic volatility model with time deformation. The latter is examined through bivariate impulse response profiles following the example of Gallant, Rossi and Tauchen (1993).

We finally study trading patterns, time deformation and stochastic volatility in foreign exchange markets. Globalization of trading in foreign exchange markets is a principal source of the daily and weekly seasonality in market volatility. One way to model such phenomena is to adopt a framework where market volatility is tied to the intensity of (world) trading through a subordinated stochastic process representation. In this essay, we combine elements from Clark (1973), Dacorogna et al. (1993) and from our work on stock prices and volume, and present a stochastic volatility model for foreign exchange markets with time deformation. The time deformation is based on daily patterns of arrival of quotes and bid-ask spreads as well as returns. For empirical estimation, we use the QMLE algorithm of Harvey et al. (1994), adapted by us for time deformed processes, and applied to the Olsen and Associates high frequency data set.

Résumé

Dans la littérature économétrique et statistique, on rencontre souvent des modèles comportant des variables latentes. Dans cette thèse, nous examinons deux catégories de modèles de ce type et nous proposons des méthodes pour les analyser. La première catégorie est un ensemble de modèles de régression où apparaissent des variables explicatives non observables telles que des variables d'anticipation. La seconde catégorie est un ensemble de modèles financiers où apparaît un processus latent de volatilité stochastique. Les problèmes de premier type sont abordés dans un premier essai tandis que nos résultats sur la seconde catégorie de modèles sont subdivisés en deux essais.

L'approche usuelle pour estimer des modèles dont les variables explicatives sont latentes ou représentent des valeurs espérées est de procéder en deux étapes. D'abord, on remplace les variables non observées par les valeurs ajustées générées à partir d'une régression auxiliaire. Ensuite, on estime le modèle structurel par les moindres carrés ordinaires. Cette méthode simple et facile à appliquer pose des problèmes au niveau des tests d'hypothèses. Notamment, les écart-types ainsi obtenus sont sous-évalués et affectent les statistiques t de sorte que les tests ne sont pas valides. Pour remédier à ce problème, diverses méthodes de correction ont été proposées. Ce sont des méthodes asymptotiques, qui le plus souvent consistent à remplacer les écart-types des moindres carrés ordinaires par les écart-types obtenus par la méthode des variables instrumentales [Pagan (1984, 1986)]. En général, les niveaux des régions de confiance résultant d'une procédure asymptotique ne sont que des approximations et peuvent parfois différer considérablement du vrai niveau à distance finie. Or, plusieurs études récentes indiquent que l'inférence basée sur les variables instrumentales en petits échantillons peut être extrêmement peu fiable notamment en présence de mauvais instruments; voir par exemple Staiger et Stock (1993), Angrist et Krueger (1994) ainsi que Dufour (1994).

Dans le premier essai nous abordons ce type de problèmes et nous dérivons deux méthodes de tests basés sur des statistiques pivotales et valides à distance finie [voir

Dufour (1994)]. La première approche que l'on appelle la "substitution d'instruments" est une généralisation de la méthode de Anderson et Rubin (1949). La seconde approche exige un partage préliminaire des données en deux sous-échantillons. Le test qu'on obtient est basé directement sur les variables explicatives générées a partir de la régression complémentaire et est plus avantageux du point de vue du nombre de dégrés de liberté conservés. Les deux types de tests sont basés sur des statistiques de type Fisher. Les régions de confiance correspondantes sont définies par des inégalités non linéaires qui, en général, peuvent être résolues soit analytiquement soit numériquement. Il faut souligner que nos méthodes d'inférence ne s'appliquent pas seulement à des modèles comportant des variables non observées. Nous les présentons dans le contexte des modèles considérés par Pagan (1986) pour la clarté d'exposition, mais nos tests peuvent également être utilisés dans divers modèles structuraux comme, par exemple, les modèles à équations simultanées.

Les distributions exactes des statistiques de tests sont obtenues sous l'hypothèse de normalité des erreurs et d'exogénéité stricte des variables explicatives du modèle. Les tests et les régions de confiance proposées demeurent valides asymptotiquement sous des hypothèses moins restrictives, permettant des erreurs non gaussiennes et des variables explicatives faiblement exogènes. Il est important de noter que les régions de confiance proposées sont non bornées avec une probabilité strictement positive [voir Dufour (1994)]. Cette caractéristique est nécessaire pour que la région soit valide pour un paramètre qui peut ne pas être identifiable et constitue une différence fondamentale entre nos méthodes et les tests du type Wald. Le niveau des régions qui ne possèdent pas cette propriété est nul. Il est clair que le problème de non identifiabilité apparaît très souvent dans l'inférence sur les paramètres de modèles qui incluent des variables non observées ou générées, ainsi que dans les modèles à équations simultanées. En simulant un modèle du premier type, on montre que les méthodes d'inférence proposées sont supérieures à la fois en termes de niveau et de puissance aux méthodes usuelles de type Wald. En outre, nous observons que la puissance des tests basés sur le partage d'échantillon s'accroît quand les variables explicatives sont estimées à partir d'une portion relativement faible de l'échantillon. Nous développons aussi une technique qui, contrairement aux tests originaux de Anderson et Rubin permet de tester les hypothèses sur un élément du vecteur de paramètres des variables endogènes ou non observées. Cette technique de projections permet aussi d'obtenir des tests sur des transformations non linéaires du vecteur de paramètres d'intérêt.

Dans la section 2 de l'essai, nous présentons le modèle principal et nous dérivons les tests et régions de confiance basés sur la méthode de substitution d'instruments. La section 3 décrit l'approche à l'inférence fondée sur un partage d'échantillon. L'étude des tests d'hypothèses conjointes sur les paramètres des variables non observées et les contraintes linéaires sur les paramètres des autres variables explicatives du modèle apparaît dans la section 4. L'extension des résultats précédents aux tests sur des composantes non anticipées des variables non observées est présentée dans la section 5. Dans la section 6 nous discutons l'inférence sur les transformations non linéaires du vecteur des paramètres d'intérêt. La validité asymptotique des tests est démontrée dans la section 7. La section 8 résume les résultats des simulations qui comparent la performance de nos tests aux autres procédures. Finalement, des applications empiriques au modèle q de Tobin et à un modèle de résultats scolaires d'élèves sont présentées dans la section 9. Le premier exemple est une illustration d'un cas où on dispose d'un ensemble d'instruments de bonne qualité, tandis que le second exemple est un modèle où, au contraire, les instruments sont faibles. Dans le premier cas les régions de confiance basées sur les statistiques de type Wald coincident avec celles qu'on obtient à partir de nos procédures. Par contre, la région de confiance obtenue dans le second exemple construite à partir de nos méthodes est non bornée. Nous concluons dans la section 10.

Dans la deuxième partie de la thèse, nous proposons un modèle de volatilité stochastique avec déformation du temps. Le modèle de volatilité stochastique connu dans la littérature décrit l'évolution des prix d'actifs financiers en temps continu, en supposant que la variance est une variable non observée qui suit une diffusion indépendante [voir, par exemple Hull et White (1987)]. En particulier, la spécification que l'on emploie dans les études empiriques est basée sur une hypothèse implicite supplémentaire d'invariance dans le temps des coefficients de deux équations. Ceci est une contrainte très forte, notamment dans le cas où on estime, par exemple, l'évolution d'un indice boursier comme le S&P500 sur une longue période. En effet, dans le passé la composition du S&P500 a été redéfinie plusieurs fois. Par ailleurs, les mécanismes de fonctionnement du marché se sont modifiés avec l'accès aux ordinateurs et le progrés en informatique.

Le modèle de volatilité stochastique exploite la prévisibilité de la volatilité à partir des variances conditionnelles passées pour déterminer les rendements d'actifs. Une approche plus ancienne explique la dynamique des rendements par le flux d'information [voir Clark (1973)]. Cette idée est justifiée par plusieurs études empiriques où l'on observe l'effet de publication de données économiques importantes sur la volatilité [voir, par exemple, Baillie et Bollerslev (1991)]. Inspirés par les deux approches, nous supposons que le processus de volatilité des prix d'actifs évolue en temps déformé (opérationnel) déterminé par le processus latent de l'arrivée d'information. Comme variables auxiliaires, on utilise des séries telles que le volume de transactions et le rendement passé. Elles servent à identifier la correspondance entre le temps calendrier et opérationnel et apparaissent dans la forme fonctionnelle de la vitesse du temps opérationnel. On y introduit aussi les valeurs absolues des rendements passés pour modéliser une asymétrie dans le comportement de la volatilité suite à un accroissement ou une baisse des prix (effet de levier). Les liens entre la dynamique des rendements financiers et celle du volume d'échanges sont analysés d'avantage dans le cadre d'un modèle bivarié de volatilité stochastique.

Contrairement aux modèles déjà existants, notre spécification admet des coefficients qui varient dans le temps et permet ainsi de mieux modéliser les divers aspects d'hétérogenéité dans la dynamique des séries financières. L'approche de déformation du temps permet aussi de distinguer une composante lisse de la volatilité sur l'échelle du temps opérationnel. En outre, l'évolution des prix est modélisé conjointement avec le volume, pour mieux identifier les périodes de haute volatilité accompagnées d'un niveau de volume accru et vice-versa.

Dans la partie empirique, nous utilisons des données journalières de la bourse de New York de 1928 à 1987. On estime les deux modèles par la méthode de quasi maximum de vraisemblance et par inférence indirecte. La raison pour employer les deux procédures est que la seconde méthode produit des estimateurs plus efficaces, mais restreint en même temps le choix de la formule du changement de temps. Dans la première approche on discrétise le modèle et dérive une représentation d'espace d'états linéaire pour maximiser ensuite la fonction de vraisemblance conditionnelle obtenue à partir d'un algorithme de filtre de Kalman. La méthode d'inférence indirecte est basée sur la fonction de densité semi-nonparametrique (SNP) estimée au préalable sur une série univariée des changements de prix et une série bivariée prix — volume. On trouve une forte évidence en faveur de l'hypothèse de présence de déformation du temps. Nous constatons que les accroissements du volume d'échanges accélèrent le temps opérationnel et rendent la volatilité moins persistante et sujette à d'importants chocs aléatoires. Les baisses des prix d'actifs ont un effet semblable, tandis que les hausses augmentent la persistance dans la volatilité et réduisent l'amplitude des chocs aléatoires en ralentissant le temps opérationnel. La complexité de la structure du modèle bivarié nous oblige à l'analyser avec des fonctions non linéaires de réponse à des impulsions (impulse responses) [voir Gallant, Rossi et Tauchen (1993)]. On y compare les effets de chocs positifs et négatifs au prix et au volume sur la volatilité et les rendements. Nos résultats coincident avec ceux de Gallant, Rossi et Tauchen obtenus à partir du modèle semi-nonparametrique bivarié.

Le modèle de base est présenté dans la section 2. L'estimation et les tests d'hypothèses sont décrits dans la section 3. La section 4 présente les résultats empiriques et la section 5 contient les conclusions.

Le troisième essai présente une extension du modèle de volatilité stochastique avec la déformation du temps. On y modélise trois séries des taux de change: DEM/US$, DEM/JPY et JPY/US$. Les prix d'offre et d'achats proviennent des enregistrements effectués par l'agence Reuter sur le marché mondial au moment de la cotation, soit en temps réel. Le marché des taux de change est actif 24 heures sur 24. L'intensité des transactions dépend pendant cette période des heures d'ouverture de centres financiers localisés en Europe, en Amérique et en Asie. Les échanges s'intensifient au moment où les marchés sur plus d'un continent deviennent accessibles, créant ainsi de forts effets “saisonniers” à travers la journée. À ces effets s'ajoutent encore les effets du jour de la semaine et les effets du mois, en déterminant conjointement une complexe composante saisonnière des rendements sur les taux de change.

L'approche de déformation de temps est très utile pour modéliser les données de haute fréquence comme on appelle les séries en temps réel. Comme les observations sont séparées par des intervalles de temps inégaux, le choix des méthodes d'estimation est très limité. Un exemple d'une adaptation du modèle ARCH à ce type de données a été présenté récemment par J. Poi et W. Polasek à la conférence sur High Frequency Data (Zürich, Mars 1995). Une autre approche consiste soit à ignorer les intervalles inégaux et à modéliser les rendements sur l'échelle du temps opérationnel déterminée par le processus de comptage des cotations, soit à échantillonner les données à des intervalles constants, comme un jour, une heure ou dix minutes. Il est crucial de prendre en considération l'ajustement du temps qui résulte de la seconde méthode. Les intervalles de temps opérationnel sont dans ce cas définies par une séquence de nombres des transactions conclues dans les intervalles du temps calendrier correspondants ou encore par le volume. Dans notre essai, nous comparons les moments empiriques des séries en temps réels et ceux produits par l'échantillonnage pour montrer les différences qui résultent d'une négligence de l'effet du changement de temps.

La forme fonctionnelle du temps opérationnel que l'on emploie dans notre modèle permet de modéliser les données échantillonnées à intervalle constant et s'adapte aux effets saisonniers. On distingue une composante prévisible du nombre des cotations dans un intervalle particulier du jour à laquelle s'ajoute une composante de surprise définie comme la différence entre la moyenne empirique et le nombre de cotations enregistrées. Une autre échelle de temps opérationnel est construite de la même façon à partir des rendements passés et des fourchettes de prix passés. Les résultats empiriques permettent de décrire la dynamique des rendements sur les taux de change dont la volatilité est, en général, particulièrement sensible aux composantes non anticipées. On compare aussi les marchés de différentes monnaies et les séries des prix d'offre et d'achat. La partie empirique est basée sur les données de Olsen and Associates du mois d'Octobre 1992 à l'Octobre 1993. Dans la section 2, nous analysons les données et nous discutons les conséquences de l'ajustement de l'échelle du temps sur les propriétés distributionnelles des données. La section 3 décrit le modèle de base. Les différentes échelles de temps opérationnel sont comparées dans la section 4. Les résultats empiriques sont présentés dans la section 5 et les conclusions sont résumées dans la section 6.

Table des matières

Liste des tableaux

Liste de figures

Finite Sample Inference Methods for Simultaneous Equations and Models with Unobserved and Generated Regressors

1. INTRODUCTION

A frequent problem in econometrics and statistics consists in making inferences on models which contain unobserved explanatory variables, such as expectational or latent variables and variables observed with error; see, for example, Barro (1977), Pagan (1984, 1986), and the recent survey of Oxley and McAleer (1993). A common solution to such problems is based on using instrumental variables which serve to replace the unobserved variables by proxies obtained from auxiliary regressions (*generated regressors*). It is also well known that the use of such regressors raises difficulties for making tests and confidence sets, and it is usually proposed to replace the ordinary least squares (OLS) standard errors by instrumental variables (IV) based standard errors; see Pagan (1984, 1986) and Murphy and Topel (1985). In any case, all the methods proposed to deal with such problems only have an asymptotic justification, which means that the resulting tests and confidence sets can be extremely unreliable in finite samples. In particular, such difficulties are especially obvious in situations which involve "weak instruments", a problem which has received considerable interest recently; see, for example, Nelson and Startz (1990a, b), Buse (1992), Maddala and Jeong (1992), Bound, Jaeger and Baker (1993, 1995), Staiger and Stock (1993), Angrist and Krueger (1994), Dufour (1994) and Hall, Rudebusch and Wilcox (1994) [for some early results relevant to the same issue, see also Nagar (1959), Richardson (1968) and Sawa (1969)].

In this paper, we treat these issues from a finite sample perspective and we propose finite sample tests and confidence sets for models with unobserved and generated regressors. We also consider a number of related problems in the more general context of linear simultaneous equations. To get reliable tests and confidence sets, we emphasize the derivation of truly pivotal (or boundedly pivotal) statistics, as opposed to statistics which are only asymptotically pivotal; for further discussion of the importance of such statistics for inference, see Dufour (1994). We study two distinct approaches for various models considered by Pagan (1984). The first one is an instrument substitution method which generalizes an approach proposed by Anderson and Rubin (1949) and Fuller (1987) for different (although related) problems, while the second one is based on splitting the sample. The instrument substitution method uses the instruments directly, instead of generated regressors, in order to test hypotheses about the "structural parameters" of interest and build confidence sets. The second approach relies on "generated regressors", which allows a gain in degrees of freedom, and a sample split technique. Depending on the problem considered, we derive either exact *similar* tests (and confidence sets) or conservative procedures. The hypotheses for which we obtain similar tests (and correspondingly similar confidence sets) include: a) hypothesis which set the value of the unobserved (expected) variable coefficient vector [as in Anderson and Rubin (1949) and Fuller (1987)]; b) analogous restrictions taken jointly with general linear constraints on the coefficients of the (observed) exogenous variables in the equation of interest; and c) hypothesis about the coefficients of "surprise" variables when such variables are included in the equation. Tests for these hypotheses are based on Fisher-type statistics, but the confidence sets typically involve nonlinear (although quite tractable) inequalities. The exact distributional theory is obtained under the assumptions of Gaussian errors and strictly exogenous regressors, which ensures that we have well-defined testable models. Although we stress here applications to models with unobserved regressors, the extensions of Anderson–Rubin procedures that we discuss are also of interest for inference in various structural models which are estimated by instrumental variable methods (e.g., simultaneous equations models). Furthermore, we show that the tests and confidence sets proposed are (locally) "asymptotically valid" under much weaker distributional assumptions (which may involve non-Gaussian errors and weakly exogenous instruments).

It is important to note that the confidence sets obtained by the methods described above, unlike Wald-type confidence sets, are unbounded with non-zero probability. As emphasized in Dufour (1994), this is a necessary property of any valid confidence set for a parameter that may not be identifiable on some subset of the parameter space. As a result, confidence procedures that do not have this property have true level zero, and the sizes of the corresponding tests (like Wald-type tests) must deviate arbitrarily from their nominal levels. It is easy to see that such difficulties occur in models with unobserved regressors, models with generated regressors, simultaneous equations models, and different types of the error-in-variables models. In the context of the first type of model, we present below simulation evidence that strikingly illustrates these difficulties. In particular, our simulation results indicate that tests based on instrument substitution methods have good power properties with respect to Wald-type tests, a feature previously pointed out for Anderson-Rubin tests by Maddala (1974) in a comparative study for simultaneous equations. Furthermore, we find that generated regressors sample-split tests perform better when the generated regressors are obtained from a relatively small fraction of the sample (e.g., 10% of the sample) while the rest of the sample is used for the main regression (in which generated regressors are used).

An apparent shortcoming of the similar procedures proposed above and what is probably one of the reasons why Anderson–Rubin tests have not become widely used, is the fact that they are restricted to testing hypotheses which specify the value of the coefficients of all the endogenous (or unobserved) explanatory variables, excluding the possibility of considering a subset of coefficients (e.g., individual coefficients). We show that inference on the individual parameters or subsets of coefficients is however feasible by applying a projection technique analogous to the ones used in Dufour (1990) and Dufour and Kiviet (1993). We also show that such techniques may be used for inference on general possibly nonlinear transformations of the parameter vector of interest.

The plan of the paper is as follows. In section 2, we describe the main model which may contain several unobserved variables (analogous to the "anticipated" parts of those variables), and we introduce the instrument substitution method for this basic model with various tests and confidence sets for the coefficients of the unobserved variables. In section 3, we propose the sample split method for the same model with again the corresponding tests and confidence sets. In section 4, we study the problem of testing joint hypotheses about the coefficients of the unobserved variables and various linear restrictions on the coefficients of other (observed) regressors in the model. Section 5 extends these results to a model which also contains error terms of the unobserved variables (the "unanticipated" parts of these variables). In section 6, we consider the problem of making inference about general nonlinear transformations of model coefficients. Finally, in section 7, we give an "asymptotic validity" result for the various procedures proposed under weaker distributional assumptions. Section 8 presents the results of simulation experiments in which the performance of our methods is compared with some widely used asymptotic procedures. Section 9 presents the empirical results on hypothesis testing in the context of the Tobin's q model and an application to an economic model from the domain of the educational economics. It explains the relationship between student's academic performance, their personal characteristics and some socio-economic factors. The first example illustrates inference in presence of good instruments, while in the second example only poor instruments are available. As expected, confidence intervals for Tobin's q based on the Wald-type procedures largely coincide with those resulting from our methods. On the contrary, large discrepancies arise between the confidence intervals obtained from the asymptotic and the exact inference methods when poor instruments are used. We conclude in section 10.

2. EXACT INFERENCE BY INSTRUMENT SUBSTITUTION

In this section we develop finite sample inference methods based on instrument substitution methods for models with unobserved and generated regressors. We first derive general formulae for the test statistics and then discuss the corresponding confidence sets.

We consider the following basic setup which includes as special cases Models 1 and 2 studied by Pagan (1984):

$$y = Z_*\delta + X\gamma + e, \tag{2.1}$$

$$Z = Z_* + V = WB + V, \tag{2.2}$$

where y is a $T \times 1$ vector of observations on a dependent variable, Z_* is a $T \times G$ matrix of unobserved variables, X is a $T \times K$ matrix of exogenous explanatory variables in the structural model, Z is a $T \times G$ matrix of observed variables, W is a $T \times q$ matrix of variables related to Z_*, while $e = (e_1, \ldots, e_T)'$ and $V = [v_1', \ldots, v_T']'$ are $T \times 1$ and $T \times G$ matrices of disturbances. The matrices of unknown coefficients, δ, γ, and B have dimensions, respectively, $G \times 1$, $K \times 1$ and $q \times G$. In order to handle common variables in both equations (2.1) and (2.2), like for example the constant term, we allow for the presence of common columns in the matrices W and X. In the setup of Pagan (1984), the exogenous regressors X are excluded from the "structural" equation (2.1).

The finite sample approach we adopt in this paper requires additional assumptions, especially on the distributional properties of the error term. We will suppose that the following conditions are satisfied:

(2.3) $$X \text{ and } W \text{ are independent of } e \text{ and } V\,;$$

(2.4) $$1 \leq \operatorname{rank}(X) = K, \quad 1 \leq \operatorname{rank}(W) = q < T, \quad 1 < K + G < T;$$

(2.5) $$(e_t,\, v_t')' \overset{\text{ind}}{\sim} N[0,\, \Omega], \quad t = 1, \ldots, T;$$

(2.6) $$\operatorname{var}(e_t) \neq 0, \quad t = 1, \ldots, T\,.$$

Assumptions (2.3) – (2.6) can be relaxed if they are replaced by assumptions on the asymptotic behaviour of the variables as $T \to \infty$. Results on the asymptotic "validity" of the various procedures proposed in this paper are presented in section 7.

Let us now consider the null hypothesis:

(2.7) $$H_0: \quad \delta = \delta_0.$$

The instrument substitution method is based on replacing the unobserved variable by a set of instruments. First, we substitute (2.2) into (2.1):

(2.8) $$y = (Z - V)\delta + X\gamma + e = Z\delta + X\gamma + (e - V\delta).$$

Then subtracting $Z\delta_0$ on both sides of (2.8), we get:

(2.9) $$\begin{aligned} y - Z\delta_0 &= Z(\delta - \delta_0) + X\gamma + (e - V\delta) \\ &= (WB + V)(\delta - \delta_0) + X\gamma + (e - V\delta) \\ &= WB(\delta - \delta_0) + X\gamma + u \end{aligned}$$

where $u = e - V\delta_0$. Now suppose that W and X have K_2 columns in common $(0 \leq K_2 < q)$ while the other columns of X are linearly independent of W:

(2.10) $$W = [W_1,\, X_2], \qquad X = [X_1,\, X_2],$$

$$\operatorname{rank}[W_1,\, X_1,\, X_2] = q + K - K_2 < T,$$

where W_1, X_1 and X_2 are $T \times q_1$, $T \times K_1$ and $T \times K_2$ matrices, respectively ($K = K_1 + K_2$, $q = q_1 + K_2$). We can then rewrite (2.9) as

$$
\begin{aligned}
(2.11) \qquad y - Z\delta_0 &= [W_1 B_1 + X_2 B_2](\delta - \delta_0) + [X_1\gamma_1 + X_2\gamma_2] + u \\
&= W_1 B_1(\delta - \delta_0) + X_1\gamma_1 + X_2[\gamma_2 + B_2(\delta - \delta_0)] + u \\
&= W_1\delta_1 + X_1\gamma_1 + X_2\gamma_{2*} + u \\
&= W_1\delta_{1*} + X\gamma_* + u
\end{aligned}
$$

where $\delta_{1*} = B_1(\delta - \delta_0)$, $\gamma_{2*} = \gamma_2 + B_2(\delta - \delta_0)$ and $\gamma_* = (\gamma_1', \gamma_{2*}')'$.

It is easy to see that model (2.11) satisfies all the assumptions of the classical linear model. Furthermore, since $\delta_{1*} = 0$ when $\delta = \delta_0$, we can test H_0 by a standard F-test of the null hypothesis

$$(2.12) \qquad H_{0*} : \delta_{1*} = 0.$$

This F-statistic has the form

$$(2.13) \qquad F(\delta_0;\, W_1) = \frac{(y - Z\delta_0)'P(M(X)W_1)(y - Z\delta_0)/q_1}{(y - Z\delta_0)'M([W_1,\, X])(y - Z\delta_0)/(T - q_1 - K)}$$

where $P(A) = A(A'A)^{-1}A'$ and $M(A) = I_T - P(A)$ for any full column rank matrix A. The statistic $F(\delta_0;\, W_1)$ may also be expressed in terms of the OLS estimator $\hat{\delta}_{1*}$ from (2.11),

$$(2.14) \qquad F(\delta_0;\, W_1) = \frac{\hat{\delta}_{1*}' W_1' M(X) W_1 \hat{\delta}_{1*}/q_1}{\hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)/(T - q_1 - K)}\,,$$

or in terms of sum of squared residuals:

$$(2.15) \qquad F(\delta_0;\, W_1) = \frac{[\hat{u}_0(\delta_0)'\hat{u}_0(\delta_0) - \hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)]/q_1}{\hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)/(T - q_1 - K)}$$

where $\hat{u}_0(\delta_0)'\hat{u}_0(\delta_0)$ and $\hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)$ denote the restricted and the unrestricted residual sum of squares from equation (2.11). When $\delta = \delta_0$, we have

(2.16) $$F(\delta_0;\, W_1) \sim F(q_1,\, T-q_1-K)$$

so that $F(\delta_0;\, W_1) \geq F(\alpha;\, q_1,\, T-q_1-K)$ is a critical region with level α for testing $\delta = \delta_0$, where

(2.17) $$\text{Prob}[F(\delta_0;\, W_1) \leq F(\alpha;\, q_1,\, T-q_1-K)] = 1-\alpha.$$

The essential ingredient of the test is the fact that $q_1 \geq 1$, i.e. some instruments must be excluded from X in (2.1). On the other hand, the usual order condition for "identification" ($q_1 \geq G$) is not necessary for applying this procedure. In other words, it is possible to test certain hypotheses about δ even if the latter vector is not completely identifiable. It is then straightforward to see that the set

(2.18) $$C_\delta(\alpha) = \{\delta_0 : F(\delta_0;\, W_1) \leq F(\alpha;\, q_1,\, T-q_1-K)\}$$

is a confidence set with level $1-\alpha$ for the coefficient δ. The tests based on the statistic $F(\delta_0;\, W_1)$ and the above confidence set generalizes the procedure proposed by Fuller (1987, pp. 16–17) for a model with one unobserved variable ($G = 1$) and X limited to a constant variable ($K = 1$).

Consider now the case where Z is a $T \times 1$ vector and X is a $T \times K$ matrix. In this case, the confidence set (2.18) for testing $H_0 : \delta = \delta_0$ has the following general form:

(2.19) $$C_\delta(\alpha) = \left\{\delta_0 : \frac{(y-Z\delta_0)'A_1(y-Z\delta_0)}{(y-Z\delta_0)'A_2(y-Z\delta_0)} \times \frac{\nu_2}{q_1} \leq F_\alpha\right\},$$

where $F_\alpha = F(\alpha;\, q_1,\, T-q_1-K)$ and $\nu_2 = T-q_1-K$ and the matrices A_1 and A_2 are defined by the expressions:

(2.20) $$A_1 = P(M(X)W_1), \quad A_2 = M([W_1,\, X]).$$

Since (ν_2/q_1) only takes positive values, the inequality in (2.19) is equivalent to

$$(y - Z\delta_0)'(A_1 - G_\alpha A_2)(y - Z\delta_0) \le 0,$$

or

$$Z'CZ\delta_0^2 - [y'CZ + Z'Cy]\delta_0 + y'Cy \le 0,$$

where $G_\alpha = (q_1/\nu_2)F_\alpha$ and $C = A_1 - G_\alpha A_2$. Adopting the usual notation, this quadratic inequality in δ_0 can be explicated as follows:

$$a\delta_0^2 + b\delta_0 + c \le 0 \tag{2.21}$$

where $a = Z'CZ$, $b = -2y'CZ$, $c = y'Cy$.

In empirical work, some problems may arise due to the high dimensions of the matrices $M(X)$ and $M([W_1, X])$. A simple way to avoid this difficulty consists in using vectors of residuals from appropriate OLS regressions. Consider the coefficient $a = Z'CZ$. We may replace it by the expression $Z'A_1Z - G_\alpha Z'A_2Z$ and then rewrite both terms as follows:

$$\begin{aligned} Z'A_1Z &= Z'P(M(X)W_1)Z \\ &= (Z'M(X))\,(M(X)W_1)\,[(M(X)W_1)'(M(X)W_1)]^{-1}(M(X)W_1)'(M(X)Z) \\ Z'A_2Z &= Z'M([W_1, X])Z = [M([W_1, X])Z]'[M([W_1, X])Z]. \end{aligned}$$

In the above expressions, $M(X)Z$ is the vector of residuals obtained by regressing Z on X, $M(X)W_1$ is the vector of residuals from the regression of W_1 on X and, finally, $M([W_1, X])Z$ is a vector of residuals from the regression of Z on X and W_1. We can proceed in the same way to compute the two other coefficients of the quadratic inequality (2.21). This will require only two additional regressions: y on X, and y on both X and W_1.

It is easy to see that the confidence set (2.21) is determined by the roots of the second order polynomial in (2.21). The shape of this confidence set depends on the signs of a and $\Delta = b^2 - 4ac$. All possible options are summarized in Table 1 where we denote by δ_{1*} the smaller root and by δ_{2*} the larger root of the second order polynomial (when both roots are real).

Table 2.1: Confidence sets based on the quadratic inequality
$a\delta_0^2 + b\delta_0 + c \le 0$

		$\Delta \ge 0$ (real roots)	$\Delta < 0$ (complex roots)
$a > 0$		$[\delta_{1*},\ \delta_{2*}]$	Empty
$a < 0$		$(-\infty,\ \delta_{1*}] \cup [\delta_{2*},\ \infty)$	$(-\infty,\ +\infty)$
$a = 0$	$b > 0$	$\left(-\infty,\ -\frac{c}{b}\right]$	
	$b < 0$	$\left[-\frac{c}{b},\ \infty\right)$	
	$b = 0,\quad c > 0$	Empty	
	$b = 0,\quad c \le 0$	$(-\infty,\ +\infty)$	

Note there are non-zero probabilities that the confidence region be empty or covers the entire real line. However, if the model is correctly specified, the probability of obtaining an empty confidence set is not greater than α. An empty confidence set might be an indication of the poor fit of the model. On the other hand, the possibility of an unbounded confidence set is a *necessary* characteristic of any valid confidence set in the present context, because the structural parameter δ may not be identifiable [see Dufour (1994)]. Unbounded confidence sets are most likely to occur when δ is not identified or close to being unidentified, for then all values of δ are observationally equivalent. Wald-type confidence sets for δ are typically bounded with probability one, so their true level must be zero. Note finally that an unbounded confidence set is not necessarily uninformative: e.g., the set $(-\infty,\ \delta_{1*}] \cup [\delta_{2*},\ \infty)$ may exclude economically important values of δ ($\delta = 0$ for example).

3. INFERENCE WITH GENERATED REGRESSORS

Test statistics similar to those of the previous section can alternatively be computed from linear regressions with generated regressors. To obtain finite sample inferences in such contexts, we propose to compute adjusted values from an independent sample. In particular, this can be done by applying a sample split technique.

Consider again the model described by (2.1) to (2.6). In (2.9), a natural thing to do would consist in replacing WB by $W\hat{B}$, where $\hat{B}$ is an estimator of B. Take $\hat{B} = (W'W)^{-1}W'Z$, the least squares estimate of B based on (2.2). Then we have:

$$(3.1) \qquad \begin{aligned} y - Z\delta_0 &= W\hat{B}(\delta - \delta_0) + X\gamma + [u + W(B - \hat{B})(\delta - \delta_0)] \\ &= \hat{Z}\delta_{0*} + X\gamma + u_* \end{aligned}$$

where $\delta_{0*} = \delta - \delta_0$ and $u_* = e - V\delta_0 + W(B - \hat{B})(\delta - \delta_0)$. Again, the null hypothesis $\delta = \delta_0$ may be tested by testing $H_{0*} : \delta_{0*} = 0$ in model (3.1). Here the standard F statistic for H_{0*} is obtained by replacing W_1 by $\hat{Z}$ in (2.13), i.e.

$$(3.2) \qquad F(\delta_0;\, \hat{Z}) = \frac{(y - Z\delta_0)'A(\hat{Z})\,(y - Z\delta_0)/G}{(y - Z\delta_0)'B(\hat{Z})\,(y - Z\delta_0)/(T - G - K)}$$

where $A(\hat{Z}) = P(M(X)\hat{Z})$, and $B(\hat{Z}) = M([\hat{Z},\, X])$. However, to get a null distribution for $F(\delta_0;\, \hat{Z})$, we will need further assumptions. In addition to the assumptions (2.1) to (2.6), we will now also suppose that

$$(3.3) \qquad e_t \text{ and } v_t \text{ are independent for each } t = 1, \ldots, T,$$

and

$$(3.4) \qquad \det(\Omega) > 0.$$

When $\delta = \delta_0 = 0$, $\hat{Z}$ and u_* are then independent and, conditional on $\hat{Z}$, model (3.1) satisfies all the assumptions of the classical linear model (with probability 1). Thus the null distribution of the statistic $F(0;\, \hat{Z})$ for testing $\delta_0 = 0$ is $F(G, T - G - K)$. Unfortunately, this property does not extend to the more general statistic $F(\delta_0;\, \hat{Z})$ where $\delta_0 \neq 0$ because $\hat{Z}$ and u_* are not independent in this case. A similar observation (in an asymptotic context) was made by Pagan (1984).

Suppose now that an estimate $\tilde{B}$ of B independent of u and V is available, and replace $\hat{Z} = W\hat{B}$ by $\tilde{Z} = W\tilde{B}$ in (3.1). We then get

$$y - Z\delta_0 = \tilde{Z}\delta_{0*} + X\gamma + u_{**} \tag{3.5}$$

where $u_{**} = e - V\delta_0 + W(B - \tilde{B})(\delta - \delta_0)$. Under the assumptions (2.1) - (2.6) and conditional on $\tilde{Z}$ (or $\tilde{B}$), model (3.5) satisfies all the assumptions of the classical linear model and

$$F(\delta_0;\, \tilde{Z}) \sim F(G,\, T - G - K) \tag{3.6}$$

when $\delta = \delta_0$, so that the critical region $F(\delta_0;\, \tilde{Z}) > F(\alpha;\, G,\, T - G - K)$ has size α. Note that the independence between u and V [assumption (3.3)] is not needed here. Furthermore (3.5) can be estimated by OLS and an alternative form of the F-statistic can be computed, i.e.

$$\begin{aligned} F(\delta_0;\, \tilde{Z}) &= \frac{\hat{\delta}_{0*}'\tilde{Z}'M(X)\tilde{Z}\hat{\delta}_{0*}/G}{\hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)/(T - G - K)} \\ &= \frac{[\hat{u}_0(\delta_0)'\hat{u}_0(\delta_0) - \hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)]/G}{\hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)/(T - G - K)}\,, \end{aligned} \tag{3.7}$$

where the usual notation has been adopted. Consequently

$$\tilde{C}_\delta(\alpha) = \{\delta_0 :\ F(\delta_0;\, \tilde{Z}) \le F(\alpha;\, G,\, T - G - K)\} \tag{3.8}$$

is a confidence set for δ with size $1-\alpha$. For scalar δ $(G = 1)$, this confidence set takes a form similar to the one in (2.19), except that $A_1 = P(M(X)\tilde{Z})$ and $A_2 = M([\tilde{Z},\, X])$.

A practical problem here consists in finding the independent estimate $\hat{B}$. Under the assumptions (2.1) - (2.6), this can be done easily by splitting the sample. Let $T = T_1 + T_2$, where $T_1 > G + K$ and $T_2 \ge q$, and write

$$y = \begin{pmatrix} y_{(1)} \\ y_{(2)} \end{pmatrix}, \quad X = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}, \quad Z = \begin{pmatrix} Z_{(1)} \\ Z_{(2)} \end{pmatrix}, \quad W = \begin{pmatrix} W_{(1)} \\ W_{(2)} \end{pmatrix} \tag{3.9}$$

where the matrices $y_{(i)}$, $X_{(i)}$, $Z_{(i)}$ and $W_{(i)}$ have T_i rows $(i = 1, 2)$. Consider now the equation

$$y_{(1)} - Z_{(1)}\delta_0 = \tilde{Z}_{(1)}\delta_{0*} + X_{(1)}\gamma + u_{(1)**} \tag{3.10}$$

where $\tilde{Z}_{(1)} = W_{(1)}\tilde{B}$ and $\tilde{B} = [W'_{(2)}W_{(2)}]^{-1}W'_{(2)}Z_{(2)}$ is obtained from the second sample. Clearly $\tilde{B}$ is independent of $\tilde{Z}_{(1)}$ and the statistic $F(\delta_0;\, \tilde{Z}_{(1)})$ based on equation (3.8) follows a $F(G,\, T_1 - K - G)$ distribution.

The sample split technique has been adopted by Angrist and Krueger (1994) to build a new IV estimator, called Split Sample Instrumental Variables (SSIV). Its advantage over the traditional IV method is that SSIV yields an estimate biased toward zero, rather than toward the probability limit of the OLS estimator in finite sample if the instruments are weak. Angrist and Krueger show that an unbiased estimate of the attenuation bias can be calculated and, consequently, an asymptotically unbiased estimator (USSIV) can be derived. In their approach, Angrist and Krueger rely on splitting the sample in half, i.e., setting $T_1 = T_2 = \frac{T}{2}$ when T is even. However, in our setup, different choices for T_1 and T_2 are clearly possible. Alternatively, one could select at random the observations assigned to the vectors $y_{(1)}$ and $y_{(2)}$. As we will show later (see section 8) the number of observations retained for the first and the second subsample have a direct impact on the power of the test. In particular, it appears that we get a more powerful test once we use a relatively small number of observations for computing the adjusted values and keep more observations for the estimation of the structural model. This point is illustrated below by simulation experiments.

4. JOINT TESTS ON δ AND γ

The instrument substitution and sample split methods described above can easily be adapted to test hypotheses on the coefficients of both the latent variables and the exogenous regressors. In this section, we derive F-type tests for general linear restrictions on the coefficient vector.

Consider again model (2.1) – (2.6), which after substituting the term $(Z - V)$ for the latent variable yields the following equation:

$$\text{(4.1)} \qquad \begin{aligned} y &= (Z - V)\delta + X\gamma + e \\ &= Z\delta + X\gamma + (e - V\delta). \end{aligned}$$

We first consider a hypothesis which fixes simultaneously δ and an arbitrary set of linear transformations of γ:

$$H_0: \ \delta = \delta_0, \qquad \nu_1 = \nu_{10}$$

where $\nu_1 = R_1\gamma$, R_1 is a $r_1 \times K$ matrix such that $1 \le \text{rank}(R_1) = r_1 \le K$.

The matrix R_1 can be viewed as a submatrix of a $K \times K$ matrix $R = [R_1', \, R_2']'$ where $\det(R) \neq 0$, so that we can write

$$R\gamma = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \gamma = \begin{bmatrix} R_1\gamma \\ R_2\gamma \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}. \tag{4.2}$$

Let $X_R = XR^{-1} = [X_{R_1}, \, X_{R_2}]$ where X_{R_1} and X_{R_2} are $T \times r_1$ and $T \times r_2$ matrices $(r_2 = K - r_1)$. Then we can rewrite (4.1) as

$$\begin{aligned} y &= Z\delta + X\gamma + (e - V\delta) \\ &= Z\delta + XR^{-1}R\gamma + (e - V\delta) \\ &= Z\delta + X_{R_1}\nu_1 + X_{R_2}\nu_2 + (e - V\delta). \end{aligned} \tag{4.3}$$

Subtracting $Z\delta_0$ and $X_{R_1}\nu_{10}$ on both sides, we get

$$\begin{aligned} y - Z\delta_0 - X_{R_1}\nu_{10} &= Z(\delta - \delta_0) + X_{R_1}(\nu_1 - \nu_{10}) + X_{R_2}\nu_2 + (e - V\delta) \\ &= [W_1B_1 + X_2B_2](\delta - \delta_0) + X_{R_1}(\nu_1 - \nu_{10}) \\ &\quad + X_{R_2}\nu_2 + [e - V\delta + V(\delta - \delta_0)]. \end{aligned} \tag{4.4}$$

Suppose now that W and X have K_2 columns in common (with $0 \le K_2 < q$), while the other columns of X are linearly independent of W as in (2.10). Since

$$X = [X_1, \, X_2] = X_R R = [X_{R_1}, \, X_{R_2}]R = [X_{R_1}, \, X_{R_2}] \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} = X_{R_1}R_1 + X_{R_2}R_2 \, ,$$

we can write

$$X = [X_1, \, X_2] = [X_{R_1}R_{11} + X_{R_2}R_{21}, \; X_{R_1}R_{12} + X_{R_2}R_{22}] \, , \tag{4.5}$$

where $R_1 = [R_{11}, \, R_{12}]$, $R_2 = [R_{21}, \, R_{22}]$ and R_{ij} is a $r_i \times K_j$ matrix $(i, j = 1, 2)$. Then replace X_2 by $X_{R_1}R_{12} + X_{R_2}R_{22}$ in (4.4):

$$(4.6) \qquad \begin{aligned} y - Z\delta_0 - X_{R_1}\nu_{10} &= [W_1B_1 + (X_{R_1}R_{12} + X_{R_2}R_{22})B_2]\,(\delta - \delta_0) \\ &\quad + X_{R_1}(\nu_1 - \nu_{10}) + X_{R_2}\nu_2 + (e - V\delta) \\ &= W_1B_1(\delta - \delta_0) + X_{R_1}[R_{12}B_2(\delta - \delta_0) \\ &\quad + (\nu_1 - \nu_{10})] + X_{R_2}\gamma_2^* + (e - V\delta) \\ &= W_1\delta_1^* + X_{R_1}\gamma_1^* + X_{R_2}\gamma_2^* + u \end{aligned}$$

where $\delta_1^* = \delta - \delta_0,\ \ \gamma_1^* = R_{12}B_2(\delta - \delta_0) + (\nu_1 - \nu_{10}),\ \ \gamma_2^* = R_{22}B_2(\delta - \delta_0) + \nu_2,$ and $u = e - V\delta$. Consequently, we can test H_0 by testing $H_0' :\ \delta_1^* = 0,\ \ \gamma_1^* = 0$ in (4.6), which leads to the statistic:

(4.7)

$$\begin{aligned} F(\delta_0,\, \nu_{10};\, W_1X_{R_1}) &= \{y\,(\delta_0,\, \nu_{10})'P(M(X_{R_2}W_{R_1})\,y\,(\delta_0,\, \nu_{10})/(q_1 + r_1)\} \\ &\quad \{y\,(\delta_0,\, \nu_{10})'M([W_1,\, X])\,y\,(\delta_0,\, \nu_{10})/(T - q_1 - K)\}^{-1} \\ &= \frac{[\hat{u}_0(\delta_0,\, \nu_{10})'\hat{u}_0(\delta_0,\, \nu_{10}) - \hat{u}_1(\delta_0,\, \nu_{10})'\hat{u}_1(\delta_0,\, \nu_{10})]/(q_1 + r_1)}{\hat{u}_1(\delta_0,\, \nu_{10})'\hat{u}_1(\delta_0,\, \nu_{10})]/(T - q_1 - K)} \end{aligned}$$

where $y\,(\delta_0,\, \nu_{10}) = y - Z\delta_0 - X_{R_1}\nu_{10},\ W_{R_1} = [W_1,\, W_{R_1}]$. $\hat{u}_0(\delta_0,\, \nu_{10})$ and $\hat{u}_1(\delta_0,\, \nu_{10})$ refer to the restricted and unrestricted residual vectors obtained from the estimation of (4.6). Under H_0, $F(\delta_0,\, \nu_{10};\, W_1,\, X_{R_1}) \sim F(q_1 + r_1,\ T - q_1 - K)$ and we reject H_0 at level α when $F(\delta_0,\, \nu_{10};\, W_1,\, X_{R_1}) \geq F(\alpha;\, q_1 + r_1,\ T - q_1 - K)$. Correspondingly,

$$(4.8) \qquad \{(\delta_0',\, \nu_{10}')' :\ F(\delta_0,\, \nu_{10};\, W_1,\, X_{R_1}) \leq F(\alpha;\, q_1 + r_1,\ T - q_1 - K)\}$$

is a confidence set with level $1 - \alpha$ for δ and $R_1\gamma_1$.

Suppose now that we employ the procedure with generated regressors and that we have calculated an estimator $\tilde{B}$ independent of u and V. We can then proceed in the following way. Since $\tilde{Z} = W\tilde{B}$, we have:

(4.9)
$$\begin{aligned}
y - Z\delta_0 - X_{R_1}\nu_{10} &= (\tilde{Z} + \tilde{V})(\delta - \delta_0) + X_{R_1}(\nu_1 - \nu_{10}) \\
&\quad + X_{R_2}\nu_2 + (e - V\delta) \\
&= \tilde{Z}(\delta - \delta_0) + X_{R_1}(\nu_1 - \nu_{10}) + X_{R_2}\nu_2 \\
&\quad + [e - V\delta + \tilde{V}(\delta - \delta_0)] \\
&= \tilde{Z}\delta_1^* + X_{R_1}\nu_1^* + X_{R_2}\nu_2 + u
\end{aligned}$$

where $\delta_1^* = \delta - \delta_0$, $\nu_1^* = \nu_1 - \nu_{10}$ and $u = e - V\delta + \tilde{V}(\delta - \delta_0)$. In this case we will simply test the hypothesis $H_0 : \delta_1^* = 0$, $\nu_1^* = 0$. The F statistic for H_0 takes the form:

(4.10)
$$\begin{aligned}
F(\delta_0, \nu_{10};\ \tilde{Z}, X_{R_1}) &= \{y(\delta_0, \nu_{10})' P(M(X_{R_2})\tilde{Z}_{R_1})\, y(\delta_0, \nu_{10})/(G + r_1)\} \\
&\quad \times \{y(\delta_0, \nu_{10})' M([\tilde{Z}, X])\, y(\delta_0, \nu_{10})/(T - G - K)\}^{-1} \\
&= \frac{[\tilde{u}_0(\delta_0, \nu_{10})'\tilde{u}_0(\delta_0, \nu_{10}) - \tilde{u}_1(\delta_0, \nu_{10})'\tilde{u}_1(\delta_0, \nu_{10})]/(G + r_1)}{\tilde{u}_1(\delta_0, \nu_{10})'\tilde{u}_1(\delta_0, \nu_{10})/(T - G - K)}
\end{aligned}$$

where $y(\delta_0, \nu_{10}) = y - Z\delta_0 - X_{R_1}\nu_{10}$, $\tilde{Z}_{R_1} = [\tilde{Z}, X_{R_1}]$, while $\tilde{u}_0(\delta_0, \nu_{10})$ and $\tilde{u}_1(\delta_0, \nu_{10})$ are the restricted and unrestricted residual vectors from (4.9). Under H_0, $F(\delta_0, \nu_{10};\ \tilde{Z}, X_{R_1}) \sim F(G + r_1,\ T - G - K)$. The corresponding critical region with level α is given by

$$F(\delta_0, \nu_{10};\ \tilde{Z}, X_{R_1}) \geq F(\alpha;\ G + r_1,\ T - G - r_1),$$

and the confidence set at level $1 - \alpha$ is thus

(4.11)
$$\{(\delta_0', \nu_{10}')' :\ F(\delta_0, \nu_{10};\ \tilde{Z}, X_{R_1}) \leq F(\alpha;\ G + r_1,\ T - G - K\}.$$

5. INFERENCE WITH A SURPRISE TERM

In many economic models we encounter the so-called surprise term among the explanatory variables. It reflects the difference between the expected value of the latent variable and the realization. In this section we study a model which contains the unanticipated part of Z [Pagan (1984, model 4)] as an additional regressor beside the latent variable, namely,

$$
\begin{aligned}
y &= Z_*\delta + (Z - Z_*)\gamma + X\beta + e \\
&= Z\delta + V\gamma + X\beta + e - V\delta\,,
\end{aligned}
\tag{5.1}
$$

$$
Z = Z_* + V = WB + V\,,
\tag{5.2}
$$

where the general assumptions (2.3) – (2.6) still hold. The term $(Z - Z_*)$ represents the unanticipated part of Z.

Consider first the problem of testing the hypothesis $H_0 : \ \delta = \delta_0$. Applying the same procedure as before, we get the equation:

$$
\begin{aligned}
y - Z\delta_0 &= Z(\delta - \delta_0) + X\beta + V\gamma + e - V\delta \\
&= WB(\delta - \delta_0) + X\beta + V\gamma + (e - V\delta_0)\,,
\end{aligned}
\tag{5.3}
$$

hence, assuming that W and X have K_2 columns in common,

$$
\begin{aligned}
y - Z\delta_0 &= [W_1B_1 + X_2B_2]\,(\delta - \delta_0) + X\beta + V(\gamma - \delta_0) + e \\
&= W_1B_1(\delta - \delta_0) + X_2B_2(\delta - \delta_0) + [X_1\beta_1 + X_2\beta_2] \\
&\quad + V(\gamma - \delta_0) + e \\
&= W_1B_1(\delta - \delta_0) + X_1\beta_1 + X_2\beta_2^* + e + V(\gamma - \delta_0) \\
&= W_1\delta_{1*} + X\beta_* + u
\end{aligned}
\tag{5.4}
$$

where $\delta_{1*} = B_1(\delta - \delta_0)$, $\beta_2^* = \beta_2 + B_2(\delta - \delta_0)$, $\beta_* = (\beta_1', \beta_2^{*\prime})'$ and $u = e + V(\gamma - \delta_0)$. Then we can test $\delta = \delta_0$ by using the F-statistic for $\delta_{10} = 0$:

$$
(5.5) \qquad F(\delta_0;\, W_1) = \frac{(y - Z\delta_0)'P(M(X)W_1)(y - Z\delta_0)/q_1}{(y - Z\delta_0)'M[X(W_1)](y - Z\delta_0)/(T - q_1 - K)}
$$

$$
= \frac{\hat{\delta}_{1*}' W_1' M(X) W_1 \hat{\delta}_{1*}/q_1}{\hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)/(T - q_1 - K)}
$$

$$
= \frac{[\hat{u}_0(\delta_0)'\hat{u}_0(\delta_0) - \hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)]/q_1}{\hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)/(T - q_1 - K)}
$$

where $\hat{u}_0(\delta_0)'\hat{u}_0(\delta_0)$ and $\hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)$ denote the corresponding restricted and unrestricted residual sum of squares from (5.4). When $\delta = \delta_0$, $F(\delta_0;\, W_1) \sim F(q_1,\, T - q_1 - K)$. It follows that $F(\delta_0;\, W_1) \geq F(\alpha;\, q - K_2,\, T - q_1 - K)$ is a critical region with level α for testing $\delta = \delta_0$ while

$$
(5.6) \qquad \{\delta_0 : \; F(\delta_0;\, W_1) \leq F(\alpha;\, q_1,\, T - q_1 - K)\}
$$

is a confidence set with level $1 - \alpha$ for δ. The procedure developed for the case where no surprise variable is present applies thus without change.

We will now generalize the previous test to a joint hypothesis about δ and any set of linear restrictions on β. The hypothesis is formulated as follows:

$$
H_0 : \; \delta = \delta_0, \quad \nu_1 = \nu_{10},
$$

where $R_1\beta = \nu_1$ and R_1 is a full line rank $r_1 \times K$ matrix. This can be done by using the corresponding F-statistic based on equation (5.4):

(5.7)

$$
F(\delta_0;\, \nu_{10}) = \frac{\begin{pmatrix} \hat{\delta}_{1*} \\ R_1\hat{\beta}_* - \nu_{10} \end{pmatrix}' [R[(W_1,\, X)'(W_1,\, X)^{-1}]R']^{-1} \begin{pmatrix} \hat{\delta}_{1*} \\ R_1\hat{\beta}_* - \nu_{10} \end{pmatrix} /(q_1 + r_1)}{\hat{u}_1(\delta_0,\, \nu_{10})'\hat{u}_1(\delta_0,\, \nu_{10})/(T - q_1 - K)}
$$

$$
= \frac{[(\hat{u}_0(\delta_0,\, \nu_{10})'\hat{u}_0(\delta_0,\, \nu_{10}) - \hat{u}_1(\delta_0,\, \nu_{10})'\hat{u}_1(\delta_0,\, \nu_{10}))]/(q_1 + r_1)}{\hat{u}_1(\delta_0,\, \nu_{10})'\hat{u}_1(\delta_0,\, \nu_{10})/(T - q_1 - K)},
$$

where $R = \begin{bmatrix} I & 0 \\ 0 & R_1 \end{bmatrix}$. Under H_0, $F\left(\binom{\delta_0}{\nu_{10}};\, W_1,\, X\right) \sim F(q_1 + r_1,\, T - q_1 - K)$, which yields the following confidence set with level $1-\alpha$ for δ and ν_1:

$$\left\{\begin{pmatrix}\delta_0\\ \nu_{10}\end{pmatrix} : F\left(\begin{pmatrix}\delta_0\\ \nu_{10}\end{pmatrix};\, W_1,\, X\right) \le F(\alpha;\, q_1 + r_1,\, T - q_1 - K)\right\}. \tag{5.8}$$

If generated regressors are used, we can write:

$$y - Z\delta_0 = W\hat{B}(\delta - \delta_0) + X\beta + V\gamma + \hat{V}(\delta - \delta). \tag{5.9}$$

Replacing $W\hat{B}$ by $W\tilde{B}$, where $\tilde{B}$ is an estimator independent of e and V, we get

$$\begin{aligned} y - Z\delta_0 &= W\tilde{B}(\delta - \delta_0) + X\beta + [e + V(\gamma - \delta_0) + \tilde{V}(\delta - \delta_0)] \\ &= \tilde{Z}\delta_* + X\beta + u \end{aligned} \tag{5.10}$$

where $\delta_* = \delta - \delta_0$ and $u = e + V(\gamma - \delta_0) + \tilde{V}(\delta - \delta_0)$. Here the hypothesis $\delta = \delta_0$ implies the null hypothesis $H_0' : \delta_* = 0$. The F-statistic for testing H_0' has the form:

$$\begin{aligned} F(\delta_0;\, \tilde{Z}) &= \frac{(y - Z\delta_0)' P(M(X)\tilde{Z})\,(y - Z\delta_0)/G}{(y - Z\delta_0)' M([\tilde{Z},\, X])\,(y - Z\delta_0)/T - G - K} \\ &= \frac{\hat{\delta}_*'\tilde{Z}'M(X)\tilde{Z}\hat{\delta}_*/G}{\hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)/(T - G - K)} \\ &= \frac{[\hat{u}_0(\delta_0)'\hat{u}_0(\delta_0) - \hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)]/G}{\hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)]/(T - G - K)} \end{aligned} \tag{5.11}$$

where, as before, $\hat{u}_0(\delta_0)$ and $\hat{u}_1(\delta_0)$ denote the restricted and unrestricted residual vectors from equation (5.10). When $\delta = \delta_0$, $F(\delta_0;\, \tilde{Z}) \sim F(G,\, T - G - K)$. Consequently, $F(\delta_0;\, \tilde{Z}) \ge F(\alpha;\, G,\, T - G - K)$ is a critical region with level α for testing $\delta = \delta_0$ and the corresponding confidence set

$$\{\delta_0 :\, F(\delta_0;\, \tilde{Z}) \le F(\alpha;\, G,\, T - G - K)\} \tag{5.12}$$

has level $1 - \alpha$. For a discussion of different forms of confidence sets when Z is a $T \times 1$ vector, see section 2.

This procedure is easily generalized again to test an hypothesis of the form $H_0 :$ $\delta = \delta_0$, $\nu_1 = \nu_{10}$, where $\nu_1 = R_1\beta$, R_1 is a $r_1 \times K$ matrix of restrictions of rank r_1,

$1 \leq r_1 \leq K$. This can be done again by estimating (5.10) with and without the restrictions $\delta = \delta_0$ and $R_1\beta = \nu_{10}$. Adopting the same notation as in the instrument substitution method, we get the F-statistic:

(5.13)

$$F(\delta_0,\, \nu_{10};\, \tilde{Z},\, X) \;=\; \frac{\begin{pmatrix} \hat{\delta}_* \\ R_1\hat{\beta} - \nu_{10} \end{pmatrix}' \left[R[(\tilde{Z},X)'(\tilde{Z},\, X)]^{-1}R'\right]^{-1} \begin{pmatrix} \hat{\delta}_* \\ R_1\hat{\beta} - \nu_{10} \end{pmatrix} / (G+r)}{\hat{u}_1(\delta_0)'\hat{u}_1(\delta_0)/(T-G-K)}$$

$$\;=\; \frac{[\hat{u}_0(\delta_0,\, \nu_{10})'\hat{u}_0(\delta_0,\, \nu_{10}) - \hat{u}_1(\delta_0,\, \nu_{10})'\hat{u}_1(\delta_0,\, \nu_{10})]/(G+r_1)}{\hat{u}_1(\delta_0,\, \nu_{10})'\hat{u}_1(\delta_0,\, \nu_{10})/(T-G-K)} \,.$$

Clearly, when $\delta = \delta_0$ and $\nu_1 = \nu_{10}$, $F\left(\binom{\delta_0}{\nu_{10}};\, \tilde{Z},\, X\right) \sim F(G+r,\, T-G-K)$, which yields the following confidence set with level $1-\alpha$ for $(\delta',\, \nu_1')'$:

$$\text{(5.14)} \qquad \left\{ \begin{pmatrix} \delta_0 \\ \nu_{10} \end{pmatrix} : F\left(\begin{pmatrix} \delta_0 \\ \nu_{10} \end{pmatrix};\, \tilde{Z},\, X \right) \leq F(\alpha;\, G+r,\, T-G-K) \right\}.$$

Let us now consider the problem of testing the hypothesis on the coefficient of the surprise term, i.e. $H_0 : \gamma = \gamma_0$. Here we will make the additional assumption:

(5.15) $\qquad e_t$ and v_t are independent for each $t = 1, \ldots, T$.

We can write

$$\begin{aligned} \text{(5.16)} \quad y &= Z_*\delta + (Z - Z_*)\gamma + X\beta + e \;=\; Z\delta + X\beta + V\gamma + e - V\delta \\ &= Z\delta + X\beta + V(\gamma - \delta) + e \;=\; (WB + V)\delta + X\beta + V(\gamma - \delta) + e \\ &= WB\delta + X\beta + V\gamma + e \;=\; WB\delta + (Z - WB)\gamma + e \\ &= WB(\delta - \gamma) + Z\gamma + X\beta + e \\ &= W_1\delta_1^* + Z\gamma + X\beta_* + e \,. \end{aligned}$$

Subtracting $Z\gamma_0$ on both sides yields

$$\begin{aligned} \text{(5.17)} \qquad y - Z\gamma_0 &= WB_* + Z(\gamma - \gamma_0) + X\beta + e \\ &= W\delta_{1*} + Z\gamma_* + X\beta_* + e \end{aligned}$$

where $\gamma_* = \gamma - \gamma_0$. We can thus test $\gamma = \gamma_0$ by testing $\gamma_* = 0$ in (5.17), which can be done by computing the F-statistic:

$$
\begin{aligned}
(5.18) \quad F(\gamma_0;\, Z) &= \frac{(y - Z\gamma_0)'P(M([W_1,\, X])Z)\,(y - Z\gamma_0)/G}{(y - Z\gamma_0)'M([W,\, Z,\, X])\,(y - Z\gamma_0)/(T - G - q_1 - K)}, \\
&= \frac{\hat{\gamma}_*'Z'M([W,\, X])Z\hat{\gamma}_*/G}{\hat{u}_1(\gamma_0)'\hat{u}_1(\gamma_0)/(T - G - q_1 - K)}, \\
&= \frac{[\hat{u}_0(\gamma_0)'\hat{u}_0(\gamma_0) - \hat{u}_1(\gamma_0)'\hat{u}_1(\gamma_0)]/G}{\hat{u}_1(\gamma_0)'\hat{u}_1(\gamma_0)/(T - G - q_1 - K)},
\end{aligned}
$$

where $\hat{u}_0(\gamma_0)$ and $\hat{u}_1(\gamma_0)$ are the restricted and unrestricted residual vectors. When $\gamma = \gamma_0$, $F(\gamma_0;\, Z) \sim F(G,\, T-G-q_1-K)$ so that $F(\gamma_0;\, Z) \geq F(\alpha;\, G,\, T-G-q_1-K)$ is a critical region with level α for $\gamma = \gamma_0$ and

$$
(5.19) \qquad \{\gamma_0 :\ F(\gamma_0;\, Z) \leq F(\alpha;\, G,\, T - G - q_1 - K)\}
$$

is a confidence set with level $1 - \alpha$ for γ.

We can now discuss the construction of the confidence set when γ is a scalar. The confidence set in (5.20) can be explicated as:

$$
(5.20) \qquad \left\{\gamma_0 :\ \frac{(y - Z\gamma_0)'D(y - Z\gamma_0)}{(y - Z\gamma_0)'E(y - Z\gamma_0)} \times \frac{\nu_2}{\nu_1} \leq F_\alpha\right\}
$$

where $\nu_1 = G = 1$, $\nu_2 = T - G - q_1 - K$, $D = P(M([W_1,\, X]))$, $E = M([W_1,\, Z,\, X])$. Define $H_\alpha = (\nu_1/\nu_2)F_\alpha$. Since the ratio ν_2/ν_1 always takes positive values, the confidence set is obtained by finding the values γ_0 that satisfy the inequality

$$
(5.21) \qquad (y - Z\gamma_0)'(D - H_\alpha E)\,(y - Z\gamma_0) \leq 0
$$

or equivalently

$$
(5.22) \qquad a\gamma_0^2 + b\gamma_0 + c \leq 0
$$

where $a = Z'LZ$, $b = -2Z'Ly$, $c = y'Ly$ and $L = D - H_\alpha E$.

Finally we can treat by similar methods the problem of testing a joint hypothesis of the type $H_0 :\ \gamma = \gamma_0,\ \nu_1 = \nu_{10}$, where $\nu_1 = R_1\beta$ and $\text{rank}(R_1) = r_1$. Again we can proceed by finding the restricted and unrestricted residual sums of squares from equation (5.13) which yields the F-statistic:

(5.23)

$$F(\gamma_0,\, \nu_{10};\, W,\, Z,\, X) \;=\; \frac{[\hat{u}_0(\gamma_0,\, \nu_{10})'\hat{u}_0(\gamma_0,\, \nu_{10}) - \hat{u}_1(\gamma_0,\, \nu_{10})'\hat{u}_1(\gamma_0,\, \nu_{10})]/(G+r)}{\hat{u}_1(\gamma_1,\, \nu_{10})'\hat{u}_1(\gamma_0,\, \nu_{10})/(T-G-q_1-K)}\,,$$

Clearly, under the null hypothesis, $F\left(\binom{\gamma_0}{\nu_{10}};\, Z,\, X\right) \sim F(G+r,\, T-G-q_1-K)$ tests and confidence sets follow as usual.

6. TESTING GENERAL NONLINEAR TRANSFORMATIONS OF δ

The finite sample tests presented in this paper are based on extensions of Anderson–Rubin statistics. As we havé pointed out in the introduction, this category of tests yields valid inferences, independently of the sample size T, performs better then other tests in terms of power and allows for unbounded confidence sets when the parameter is unidentified or nearly unidentified. However, an apparent limitation of Anderson–Rubin type tests come from the fact that they are designed for hypothesis fixing the complete vector of the of the endogenous (or unobserved) regressor coefficients. In this section, we propose a solution to this problem which is based on applying a projection technique. Even more generally, we study inference on general nonlinear transformations of δ in (2.1) and propose finite sample tests of general restrictions on subvectors of δ. For a similar approach, see Dufour (1989, 1990) and Dufour and Kiviet (1994).

In the previous sections, we derived confidence sets for δ which take the form

(6.1) $$C_\delta(\alpha) = \{\delta_0 :\; F(\delta_0;\, W_1) \le F(\alpha)\}.$$

This means the probability that the true δ belongs to the set $C_\delta(\alpha)$ is:

(6.2) $$P[\delta \in C_\delta(\alpha)] = 1-\alpha.$$

If $\delta = \delta_0$, we have

$$P[\delta_0 \in C_\delta(\alpha)] = 1 - \alpha\,, \quad \text{and} \quad P[\delta_0 \notin C_\delta(\alpha)] = \alpha. \tag{6.3}$$

Consider first a nonlinear transformation of δ: $\eta = f(\delta)$. It is easy to see that the set

$$C_\eta(\alpha) = \{\eta_0 : \ \eta_0 = f(\delta) \ \text{ for some } \ \delta \in C_\delta(\alpha)\} \tag{6.4}$$

is a confidence set for η with level at least $1 - \alpha$: i.e.

$$P[\eta \in C_\eta(\alpha)] \geq 1 - \alpha \tag{6.5}$$

hence,

$$P[\eta \notin C_\eta(\alpha)] \leq \alpha. \tag{6.6}$$

Consequently, by rejecting $H_0 : \eta = \eta_0$ whenever $\eta_0 \notin C_\eta(\alpha)$, we perform a test of level α. Furthermore

$$\eta_0 \notin C_\eta(\alpha) \Leftrightarrow \eta_0 \neq f(\delta_0)\,, \quad \forall\, \delta_0 \in C_\delta(\alpha) \tag{6.7}$$

so that the condition $\eta_0 \notin C_\eta(\alpha)$ can be verified by minimizing $F(\delta_0; W_1)$ over the set $f^{-1}(\eta_0) = \{\delta_0 : \ f(\delta_0) = \eta_0\}$.

Consider now the special case where $\eta = f(\delta) = \delta_1$ and $\delta = (\delta_1', \delta_2')'$, i.e. η is a subvector of δ. Then the confidence set $C_\eta(\alpha)$ takes the form:

$$C_\eta(\alpha) = C_{\delta_1}(\alpha) = \left\{\delta_{10} : \begin{pmatrix}\delta_{10} \\ \delta_2\end{pmatrix} \in C_\delta(\alpha), \text{ for some } \delta_2\right\}. \tag{6.8}$$

Consequently we must have:

$$P[\delta_1 \in C_{\delta_1}(\alpha)] \geq 1 - \alpha \tag{6.9}$$

and

$$P[\delta_{10} \notin C_{\delta_1}(\alpha)] \leq \alpha. \tag{6.10}$$

The test which rejects $H_0 : \delta_1 = \delta_{10}$ when $\delta_{10} \notin C_{\delta_1}(\alpha)$ has level not greater than α. Furthermore,

$$(6.11) \qquad \delta_{10} \notin C_{\delta_1}(\alpha) \Leftrightarrow \forall\, \delta_2, \quad F\left(\begin{pmatrix}\delta_{10}\\ \delta_2\end{pmatrix};\, W_1\right) \geq F(\alpha)$$

where the last equivalence holds with probability one. In practice, we can employ a numerical procedure which would consist in minimizing the F statistic with respect to δ_2 and comparing the minimum value of F with $F(\alpha)$. In case where the minimum of F is greater than $F(\alpha)$ in which case we reject the hypothesis $\delta_1 = \delta_{10}$.

7. ASYMPTOTIC VALIDITY

In this section we show that the finite sample inference methods remain valid under weaker assumptions when the number of observations goes to infinity.

Consider again the model described by (2.1) – (2.6) and (2.10), which yields the following equations:

$$(7.1) \qquad y = Z\delta + X\gamma + u$$

$$(7.2) \qquad Z = W_1 B_1 + X_2 B_2 + V,$$

where $u = e - V\delta$. If we are prepared to accept a procedure which is only asymptotically "valid", we can relax the finite-sample assumptions (2.3) – (2.6) since the normality of error terms and their independence are no longer necessary. We will prove the asymptotic validity of our tests by showing that:

a) under the null hypothesis $\delta = \delta_0$ the F-statistic in (2.13),

$$F(\delta_0;\, W_1) = \frac{(y - Z\delta_0)' M(X) W_1 [W_1' M(X) W_1]^{-1} W_1' M(X)\,(y - Z\delta_0)/q_1}{(y - Z\delta_0)' M([X,\, W_1])\,(y - Z\delta_0)/(T - q - K)}$$

follows a $\chi^2_{q_1}/q_1$ distribution asymptotically (as $T \to \infty$), where $q_1 = q_1 - K$;

b) under the fixed alternative $\delta = \delta_1$, provided that $B_1(\delta_1 - \delta_0) \neq 0$, the value of (2.13) tends to get infinitely large as T increases, i.e. the test based on $F(\delta_0;\, W_1)$ is consistent.

Assume that the following limits hold jointly:

1) $\left(\frac{u'u}{T},\ \frac{u'V}{T},\ \frac{V'V}{T}\right) \rightarrow (\sigma_u^2,\ \Sigma_{uV},\ \Sigma_V)$,

2) $\left(\frac{X'X}{T},\ \frac{X'W_1}{T},\ \frac{W_1'W_1}{T}\right) \rightarrow (\Sigma_{XX},\ \Sigma_{XW_1},\ \Sigma_{W_1W_1})$,

3) $\left(T^{-\frac{1}{2}}X'u,\ T^{-\frac{1}{2}}W_1'u,\ T^{-\frac{1}{2}}X'V,\ T^{-\frac{1}{2}}W_1'V,\right)$
$\Rightarrow\ \Phi\ \equiv\ (\Phi_{Xu},\ \Phi_{W_1u},\ \Phi_{XV},\ \Phi_{W_1V})$,

where $\rightarrow$ and $\Rightarrow$ denote respectively convergence in probability and convergence in distribution as $T \rightarrow \infty$, and the joint distribution of the random variables in Φ is multinormal with the covariance matrix of $(\Phi'_{Xu},\ \Phi'_{W_1u})'$ given by

$$\Sigma = V\begin{bmatrix}\Phi_{Xu}\\ \Phi_{W_1u}\end{bmatrix} = \begin{bmatrix}\Sigma_{XX} & \Sigma_{XW_1}\\ \Sigma_{XW_1} & \Sigma_{W_1W_1}\end{bmatrix}$$

where $\Sigma_{XW_1} = \Sigma'_{XW_1}$ and $\det(\Sigma) \neq 0$. We know from equation (2.11) that

$$y - Z\delta_0 = W_1B_1(\delta - \delta_0) + X\gamma_* + u.$$

Under the null hypothesis $\delta = \delta_0$, the numerator of (2.13) is equal to N where N is defined by the expression:

$$\begin{aligned} N &= u'M(X)W_1[W_1'M(X)W_1]^{-1}W_1'M(X)u/q_1 \\ &= u'(I-P)W_1[W_1'(I-P)W_1]^{-1}W_1'(I-P)u/q_1 \\ &= \left[T^{-\frac{1}{2}}W_1'(I-P)u\right]'\left[\frac{1}{T}W_1'(I-P)W_1\right]^{-1}\left[T^{-\frac{1}{2}}W_1'(I-P)u\right]/q_1 \end{aligned}$$

where $P = P(X) = X(X'X)^{-1}X'$. Under the assumptions 1 to 3, we have the following convergence result:

$$\begin{aligned} T^{-\frac{1}{2}}W_1'(I-P)u &= T^{-\frac{1}{2}}W_1'u - \left(\frac{1}{T}W_1'X\right)\left(\frac{1}{T}X'X\right)^{-1}\left(T^{-\frac{1}{2}}X'u\right) \\ &\Rightarrow\ \Phi_{W_1|X} \equiv \Phi_{W_1u} - \Sigma_{W_1X}\,\Sigma_{XX}^{-1}\,\Phi_{Xu} \end{aligned}$$

where

$$
\begin{aligned}
V[\Phi_{W_1|X}] &= V[\Phi_{W_1u}] + \Sigma_{W_1X}\,\Sigma_{XX}^{-1}\,V[\Phi_{Xu}]\,\Sigma_{XX}^{-1}\,\Sigma_{XW_1} \\
&\quad - E[\Phi_{W_1u}\,\Phi'_{Xu}]\,\Sigma_{XX}^{-1}\,\Sigma_{XW_1} - \Sigma_{W_1X}\,\Sigma_{XX}^{-1}\,E[\Phi_{Xu}\,\Phi'_{W_1u}] \\
&= \Sigma_{W_1W_1} + \Sigma_{W_1X}\,\Sigma_{XX}^{-1}\,\Sigma_{XX}\,\Sigma_{XX}^{-1}\,\Sigma_{XW_1} - 2\Sigma_{W_1X}\,\Sigma_{XX}^{-1}\,\Sigma_{XW_1} \\
&= \Sigma_{W_1W_1} - \Sigma_{W_1X}\,\Sigma_{XX}^{-1}\,\Sigma_{XW_1}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{T}\,W_1'(I-P)W_1 &= \frac{1}{T}W_1'W_1 - \frac{1}{T}\,W_1'X\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{T}\,X'W_1\right) \\
&\to \Sigma_{W_1W_1} - \Sigma_{W_1X}\,\Sigma_{XX}^{-1}\,\Sigma_{XW_1}.
\end{aligned}
$$

Consequently

$$
N \Rightarrow \Phi'_{W_1|X}\left(\Sigma_{W_1W_1} - \Sigma_{W_1X}\,\Sigma_{XX}^{-1}\,\Sigma'_{XW_1}\right)^{-1}\Phi_{W_1|X}/q_1 \sim \chi^2(q_1)/q_1.
$$

This means that we can define the confidence intervals as the sets of points $\{\delta_0\}$ for which the statistic (2.13) fails to reject, using the asymptotic $\chi^2_{q_1}/q_1$ critical values or the more conservative critical values of the Fisher distribution. Furthermore, it is easy to see that, both under the null and the alternative, the denominator D converges to σ_u^2:

$$
\begin{aligned}
D &= u'M([X,\,W_1])u/T \\
&= \frac{u'u}{T} - \frac{u'[X,\,W_1]\{[X,\,W_1]'[X,\,W_1]\}^{-1}[X,\,W_1]u}{T} \to \sigma_u^2.
\end{aligned}
$$

Consider now a fixed alternative $\delta = \delta_1$. We need to show that N goes to infinity as T gets large. When $\delta = \delta_0$, we have

$$N = [W_1B_1(\delta_1-\delta_0)+u]'M(X)W_1[W_1'M(X)W_1]^{-1}W_1'M(X)\,[W_1B_1(\delta_1-\delta_0)+u]/q_1$$

$$= \left[T^{-\frac{1}{2}}(W_1'M(X)W_1B_1(\delta_1-\delta_0)+W_1'M(X)u)\right]'\left[\frac{W_1'M(X)W_1}{T}\right]^{-1}$$

$$\left[T^{-\frac{1}{2}}(W_1'M(X)W_1B_1(\delta_1-\delta_0)+W_1'M(X)u)\right]/q_1.$$

The behavior of the variable N depends on the convergence limits of the terms on the right-hand side of the last equation. It means that we can find the limit of N by showing the convergence of the individual components. The major building block of the expression for N is

$$T^{-\frac{1}{2}}[W_1'M(X)W_1B_1(\delta_1-\delta_0)+W_1'M(X)u] = T^{\frac{1}{2}}\left(\frac{W_1'M(X)W_1}{T}\right)B_1(\delta_1-\delta_0) + T^{-\frac{1}{2}}W_1'M(X)u.$$

As we have shown, $T^{-\frac{1}{2}}W_1'M(X)u$ converges in distribution to a random variable $\Phi_{W_1|X}$ and the term $T^{\frac{1}{2}}\left(\frac{W_1'M(X)W_1}{T}\right)B_1(\delta_1-\delta_0)$ diverges in probability as T gets large. Consequently, under a fixed alternative, the whole expression goes to infinity, and the test is consistent. It is easy to prove similar asymptotic results for the other tests proposed in this paper.

8. MONTE CARLO STUDY

In this section, we present the results of a small Monte Carlo experiment comparing the performance of the exact tests proposed above with other available (asymptotically justified) procedures, especially Wald-type procedures.

A total number of one thousand realizations of an elementary version of the model (2.1) – (2.2), equivalent to the Model 1 discussed by Pagan (Pagan, 1984) have been simulated for a sample of size $T = 100$. In this particular specification, only one latent variable Z is present. The error terms in e and V (where e and V are vectors of length 100) are independent with $N(0, 1)$ distributions. We allow for the presence of only one instrumental variable W in the simulated model, which was also independently drawn from a $N(0, 1)$ distribution. According to the Pagan's original specification, there is no constant term or any exogenous variables included.

The explanatory power of the instrumental variable W depends on the value of the parameter B. Hence, we let B vary and take, respectively, the following values: 0, 0.05, 0.1, 0.5 and 1. Note that when B is equal or very close to zero, W has almost no explanatory power, i.e. W is a bad instrument for the latent variable Z. For each value of B we consider four null hypotheses

$$H_0: \ \delta = \delta_0, \ \text{ for } \ \delta_0 = 0, \ 1, \ 5, \ 10 \ \text{ and } \ 50,$$

each one being tested against four alternative hypotheses of the form

$$H_1: \ \delta = \delta_1 \ , \quad \text{for} \quad \delta_1 = \delta_0 + p^* I(\delta_0) \ .$$

The alternative H_1 is constructed by adding an increment to the value of δ_0 where $p^* = 0$, 0.5, 1, 2 and 4, and $I(\delta_0) = 1$ for $\delta_0 = 0$, and $I(\delta_0) = \delta_0$ otherwise.

Table 1 summarizes the results. In the first 3 columns, we report the parameter value of B, δ_0 and the alternative δ_1. When the entries in columns II and III are equal, we have $\delta_0 = \delta_1$, and the corresponding row reports the levels of the tests. The next three columns (IV, V and VI) show the performance of the Wald-type IV-based test [as proposed by Pagan (1984)], which consists in correcting the understated standard errors of a two stage procedure by replacing them by a 2SLS standard error. We report the corresponding results in column IV. In cases where the level of Pagan's test exceeds 5%, we consider two correction methods. The first method (column V) is based on the critical value of the test at a 5% level for a given value of δ_0 in the particular row of the table. The critical value is obtained from an independent simulation with 1000 realizations of the model. Another independent simulation allows us to compute the critical value at 5% level in an extreme case when the instrumental variable is very bad, i.e. by supposing $B = 0$ also for a given value of δ_0

in the given row (column VI). This turns out to yield larger critical values and is thus closer to the theoretically correct critical value to be used here (on the assumption that B is actually unknown). In column VII, we present the power of the exact test based on the instrument substitution. In the following four columns, (VIII to XI), we show the performance of the exact test based on splitting the sample, where the numbers of observations used to estimate the structural equation are, respectively, 25, 50, 75 and 90. Finally, we report the level and the power of a naive two-stage test as well as the results of a test obtained by replacing the latent variable Z_* in the structural equation by the observed values Z.

Let us first discuss the reliability of the asymptotic procedures. The level of the IV test proposed by Pagan exceeds 5% essentially always when the parameter B is less then 0.5, sometimes by very wide margins. The tests based on the two-stage procedure or replacing the latent variable by the vector of observed values are both extremely unreliable no matter the value of the parameter B.

The performance of Pagan's test improves once we move to higher values of the parameter B, i.e. when the quality of the instrument increases. The improvement is observed both in terms of the level and power. It is, however, important to note that Pagan's test has, in general, the same or less power, than the exact tests. The only exception is the sample split test reported in column 8, where only 25 observations were retained to estimate the structural equation.

For B higher then 0.5, the two other asymptotic tests are still performing worse then the other tests. They are indeed extremely unreliable. In the same range of B, the exact tests behave very well. They show the best power properties compared to the asymptotically based procedures and in general outperform the other tests.

Table 8.1: SIMULATION BASED STUDY OF TEST PERFORMANCE FOR A MODEL WITH UNOBSERVED VARIABLES

Parameter Values			Rejection Frequencies									
B	δ_0	δ_1	Pagan-type			IS	Split-sample				2S	OLS
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
0.00	0.0	0.0	0.1	.	.	5.1	5.1	6.1	5.2	5.4	5.1	.
0.00	0.0	0.5	0.0	.	.	4.7	5.1	4.4	4.1	3.9	4.7	.
0.00	0.0	1.0	0.0	.	.	5.6	4.8	5.5	5.7	5.4	5.6	.
0.00	0.0	2.0	0.0	.	.	4.2	4.5	4.5	3.8	4.5	4.2	.
0.00	0.0	4.0	0.0	.	.	5.2	5.3	5.9	4.3	5.0	5.2	.
0.00	1.0	1.0	7.3	5.1	5.1	5.0	4.6	4.9	4.8	5.2	15.7	4.7
0.00	1.0	1.5	6.8	5.5	5.5	4.4	4.8	4.4	5.4	6.1	15.7	6.8
0.00	1.0	2.0	7.6	5.9	5.9	5.0	4.3	4.8	4.8	5.1	17.9	6.5
0.00	1.0	3.0	8.6	6.6	6.6	6.3	5.0	4.9	5.0	5.8	19.9	7.0
0.00	1.0	5.0	6.6	4.9	4.9	4.4	4.3	4.6	5.5	4.6	18.1	5.1
0.00	5.0	5.0	54.1	5.5	5.5	5.1	5.5	4.2	5.2	4.9	70.5	69.3
0.00	5.0	7.5	52.8	5.4	5.4	4.9	6.1	4.9	5.1	4.6	69.7	69.0
0.00	5.0	10.0	56.5	5.7	5.7	4.8	4.5	6.1	5.0	4.8	71.7	71.5
0.00	5.0	15.0	50.7	4.6	4.6	4.8	4.5	4.3	4.5	3.8	66.6	67.0
0.00	5.0	25.0	52.7	5.2	5.2	4.6	4.5	4.6	5.6	5.0	67.8	68.8
0.00	10.0	10.0	69.0	4.5	4.5	4.9	5.3	6.0	4.9	5.1	84.5	85.0
0.00	10.0	15.0	68.4	5.7	5.7	5.9	4.7	5.0	5.6	4.5	84.3	83.9
0.00	10.0	20.0	68.6	5.0	5.0	5.7	4.3	4.9	4.7	5.2	84.6	84.3
0.00	10.0	30.0	70.2	4.9	4.9	4.5	5.4	5.2	5.0	5.2	85.4	84.4
0.00	10.0	50.0	68.7	5.3	5.3	4.8	4.2	5.1	5.6	5.0	83.6	83.1
0.00	50.0	50.0	86.5	6.4	6.4	5.4	4.4	5.0	5.1	5.4	96.9	96.5
0.00	50.0	75.0	85.2	6.7	6.7	6.2	3.9	5.0	6.6	6.7	95.1	96.1
0.00	50.0	100.0	87.4	5.2	5.2	4.6	6.5	5.0	4.5	5.5	96.8	96.4
0.00	50.0	150.0	85.8	6.5	6.5	5.8	5.0	5.3	5.9	5.9	97.1	97.1
0.00	50.0	250.0	86.7	6.8	6.8	5.9	4.8	6.0	6.2	5.8	97.1	97.3
0.05	0.0	0.0	0.0	.	.	4.8	5.0	3.6	3.6	5.3	4.8	.
0.05	0.0	0.5	0.2	.	.	4.9	5.1	5.5	4.8	5.2	4.9	.
0.05	0.0	1.0	0.0	.	.	7.4	5.4	5.7	6.2	7.6	7.4	.
0.05	0.0	2.0	0.3	.	.	16.6	8.7	11.7	14.7	15.7	16.6	.
0.05	0.0	4.0	1.0	.	.	47.8	16.4	26.9	38.1	44.0	47.8	.

Table 8.1 cont.

0.05	1.0	1.0	6.9	5.2	5.6	4.7	4.8	4.4	4.8	5.5	16.9	7.9
0.05	1.0	1.5	6.0	4.6	4.7	5.4	6.0	6.0	5.4	5.2	16.9	7.5
0.05	1.0	2.0	4.7	3.9	3.9	5.3	5.7	4.6	5.1	5.2	18.1	7.6
0.05	1.0	3.0	4.0	2.7	2.7	9.9	6.3	7.4	8.4	10.5	25.3	7.4
0.05	1.0	5.0	2.6	2.1	2.1	27.0	9.0	14.9	23.2	25.4	51.1	5.6
0.05	5.0	5.0	33.8	4.6	1.6	4.6	5.8	5.3	5.2	4.8	71.7	72.7
0.05	5.0	7.5	21.0	2.3	0.2	6.3	4.8	4.6	5.3	6.0	69.7	71.4
0.05	5.0	10.0	12.4	0.4	0.1	8.7	4.8	5.6	7.6	8.5	71.9	69.9
0.05	5.0	15.0	5.1	0.1	0.0	14.8	6.1	8.6	11.7	13.2	81.2	66.9
0.05	5.0	25.0	3.9	0.0	0.0	47.1	15.3	26.2	39.1	43.0	93.6	59.0
0.05	10.0	10.0	34.9	7.6	0.2	6.3	6.6	6.3	6.4	6.5	84.8	84.0
0.05	10.0	15.0	22.9	1.3	0.0	6.4	4.4	5.8	5.8	5.9	85.8	78.9
0.05	10.0	20.0	14.1	0.6	0.0	8.6	5.1	6.1	6.7	7.6	88.9	79.0
0.05	10.0	30.0	5.1	0.0	0.0	14.5	6.7	10.4	13.3	13.9	90.0	74.2
0.05	10.0	50.0	4.4	0.1	0.0	52.5	18.6	30.1	40.8	49.1	97.5	62.2
0.05	50.0	50.0	32.7	5.1	0.0	4.7	4.7	6.0	5.2	4.5	97.5	92.0
0.05	50.0	75.0	21.2	1.7	0.0	6.4	4.5	4.9	5.3	6.2	96.9	89.2
0.05	50.0	100.0	14.3	0.6	0.0	8.5	5.8	7.0	7.2	7.3	97.7	86.5
0.05	50.0	150.0	6.4	0.3	0.0	17.6	7.0	11.1	15.1	15.8	97.0	79.8
0.05	50.0	250.0	3.2	0.0	0.0	51.3	16.0	28.3	38.7	46.1	99.8	65.3
0.10	0.0	0.0	0.0	.	.	4.8	4.2	4.9	4.5	5.0	4.8	.
0.10	0.0	0.5	0.2	.	.	8.2	6.8	7.1	6.9	7.4	8.2	.
0.10	0.0	1.0	0.1	.	.	15.8	7.1	8.9	13.9	13.5	15.8	.
0.10	0.0	2.0	2.4	.	.	49.4	16.9	29.3	40.7	46.0	49.4	.
0.10	0.0	4.0	8.8	.	.	97.1	47.7	78.9	93.2	95.9	97.1	.
0.10	1.0	1.0	7.3	4.4	5.6	4.7	5.3	5.1	4.5	4.7	15.2	14.0
0.10	1.0	1.5	4.4	2.9	3.8	6.6	4.4	5.6	6.3	6.2	19.8	16.2
0.10	1.0	2.0	3.0	1.9	2.3	10.6	6.6	7.3	9.5	10.0	25.8	14.3
0.10	1.0	3.0	0.9	0.7	0.9	28.3	9.3	18.7	23.8	26.6	49.5	10.9
0.10	1.0	5.0	0.6	0.3	0.5	80.1	26.4	49.4	66.1	74.1	92.4	7.4
0.10	5.0	5.0	17.4	4.6	0.6	5.2	5.2	4.7	4.8	5.4	71.5	78.9
0.10	5.0	7.5	5.8	1.1	0.0	7.2	6.0	6.4	7.4	7.5	73.7	74.4
0.10	5.0	10.0	2.3	0.2	0.0	16.5	7.9	11.1	14.0	16.0	81.6	73.0
0.10	5.0	15.0	1.0	0.0	0.0	50.5	15.4	27.2	38.7	45.7	94.8	65.2
0.10	5.0	25.0	0.4	0.0	0.0	97.0	45.5	76.6	89.4	95.0	100.0	46.9

Table 8.1 cont.

0.10	10.0	10.0	17.1	5.6	0.0	4.7	4.6	4.7	6.0	5.7	84.6	86.0
0.10	10.0	15.0	6.0	1.5	0.0	7.0	6.4	7.0	8.0	6.7	85.0	84.8
0.10	10.0	20.0	2.7	0.1	0.0	14.1	6.5	10.4	11.3	13.2	90.7	79.4
0.10	10.0	30.0	0.8	0.0	0.0	51.9	18.0	28.8	40.9	47.9	97.8	68.9
0.10	10.0	50.0	0.5	0.1	0.0	96.5	49.5	77.6	91.6	94.1	100.0	49.3
0.10	50.0	50.0	19.8	4.8	0.0	5.9	4.5	5.1	5.1	4.8	97.0	89.6
0.10	50.0	75.0	6.5	0.8	0.0	7.7	5.5	5.7	6.6	6.6	97.4	86.1
0.10	50.0	100.0	3.5	0.5	0.0	17.7	9.4	12.3	15.7	17.3	97.7	82.2
0.10	50.0	150.0	0.9	0.0	0.0	45.9	16.4	27.7	39.5	43.5	99.6	73.1
0.10	50.0	250.0	0.8	0.0	0.0	97.2	48.9	78.5	94.0	95.6	100.0	49.7
0.50	0.0	0.0	2.7	.	.	4.6	5.4	4.3	4.8	4.4	4.6	.
0.50	0.0	0.5	60.3	.	.	67.7	24.1	41.8	55.0	63.8	67.7	.
0.50	0.0	1.0	98.8	.	.	99.9	68.7	92.8	99.1	99.6	99.9	.
0.50	0.0	2.0	99.6	.	.	100.0	98.4	100.0	100.0	100.0	100.0	.
0.50	0.0	4.0	99.0	.	.	100.0	100.0	100.0	100.0	100.0	100.0	.
0.50	1.0	1.0	5.3	4.8	4.2	5.0	4.7	5.1	4.9	4.6	17.6	98.4
0.50	1.0	1.5	8.5	5.2	2.6	41.4	15.5	24.4	32.4	39.3	64.4	92.8
0.50	1.0	2.0	68.0	58.1	47.4	93.4	39.7	68.6	84.3	90.6	98.4	62.6
0.50	1.0	3.0	98.7	98.2	97.5	100.0	90.3	99.8	100.0	100.0	100.0	1.7
0.50	1.0	5.0	99.8	99.7	99.6	100.0	100.0	100.0	100.0	100.0	100.0	0.1
0.50	5.0	5.0	7.4	5.6	0.0	5.1	4.2	5.0	4.4	5.3	69.6	100.0
0.50	5.0	7.5	9.7	1.7	0.0	66.6	18.4	39.4	54.5	61.6	97.7	99.9
0.50	5.0	10.0	92.6	69.1	0.0	99.7	63.9	90.5	97.9	99.4	100.0	99.2
0.50	5.0	15.0	99.1	97.9	0.0	100.0	98.8	100.0	100.0	100.0	100.0	5.4
0.50	5.0	25.0	99.6	99.1	0.0	100.0	100.0	100.0	100.0	100.0	100.0	0.1
0.50	10.0	10.0	6.9	5.2	0.0	5.1	5.5	5.2	4.2	5.6	83.5	100.0
0.50	10.0	15.0	8.6	1.0	0.0	67.9	21.7	39.9	55.4	62.0	99.6	99.7
0.50	10.0	20.0	92.1	74.2	0.0	99.7	66.6	93.2	98.7	99.8	100.0	99.1
0.50	10.0	30.0	99.5	99.0	0.0	100.0	99.4	100.0	100.0	100.0	100.0	5.6
0.50	10.0	50.0	99.5	99.1	0.0	100.0	100.0	100.0	100.0	100.0	100.0	0.0
0.50	50.0	50.0	8.3	6.7	0.0	4.6	3.9	4.5	4.4	4.5	96.3	100.0
0.50	50.0	75.0	8.9	3.7	0.0	69.8	21.8	39.1	56.1	64.7	99.9	100.0
0.50	50.0	100.0	94.3	88.8	0.0	99.6	63.2	92.3	98.5	99.5	100.0	99.4
0.50	50.0	150.0	98.8	98.3	0.0	100.0	99.4	100.0	100.0	100.0	100.0	5.2
0.50	50.0	250.0	99.5	99.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	0.3

Table 8.1 cont.

1.00	0.0	0.0	5.1	.	.	5.6	4.9	5.0	5.6	5.8	5.6	.
1.00	0.0	0.5	99.5	.	.	99.5	64.9	91.2	98.5	99.2	99.5	.
1.00	0.0	1.0	100.0	.	.	100.0	99.2	100.0	100.0	100.0	100.0	.
1.00	0.0	2.0	100.0	.	.	100.0	100.0	100.0	100.0	100.0	100.0	.
1.00	0.0	4.0	100.0	.	.	100.0	100.0	100.0	100.0	100.0	100.0	.
1.00	1.0	1.0	6.8	7.2	3.8	6.3	5.4	7.0	6.9	6.8	17.9	99.7
1.00	1.0	1.5	87.9	89.2	82.2	93.3	39.5	68.3	84.7	90.1	98.1	33.7
1.00	1.0	2.0	100.0	100.0	100.0	100.0	89.9	99.8	100.0	100.0	100.0	0.7
1.00	1.0	3.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	57.3
1.00	1.0	5.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.1
1.00	5.0	5.0	4.8	4.4	0.0	4.1	5.5	4.4	4.7	4.8	67.2	100.0
1.00	5.0	7.5	98.8	98.3	0.0	99.6	62.5	91.5	98.0	99.4	100.0	67.6
1.00	5.0	10.0	100.0	100.0	0.0	100.0	99.0	100.0	100.0	100.0	100.0	1.3
1.00	5.0	15.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	65.9
1.00	5.0	25.0	100.0	100.0	7.3	100.0	100.0	100.0	100.0	100.0	100.0	98.3
1.00	10.0	10.0	5.1	4.4	0.0	6.0	6.2	5.8	6.9	6.3	85.3	100.0
1.00	10.0	15.0	98.8	98.5	0.0	99.6	63.1	91.1	97.7	99.4	100.0	69.5
1.00	10.0	20.0	100.0	100.0	0.0	100.0	99.0	100.0	100.0	100.0	100.0	0.6
1.00	10.0	30.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	66.5
1.00	10.0	50.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	99.2
1.00	50.0	50.0	5.2	5.0	0.0	5.5	5.5	5.3	5.2	6.9	96.8	100.0
1.00	50.0	75.0	99.0	98.7	0.0	99.9	65.8	91.4	98.3	99.3	100.0	68.1
1.00	50.0	100.0	100.0	100.0	0.0	100.0	98.8	100.0	100.0	100.0	100.0	0.6
1.00	50.0	150.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	67.0
1.00	50.0	250.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	99.0

Notes:

Column:

- I: value of parameter B;
- II: null hypothesis;
- III: alternative hypothesis;
- IV: Pagan's test;
- V: Pagan's test corrected for level;
- VI: Pagan's test corrected for level at $B = 0$;
- VII: instrument substitution test (IS);

Column:

- VIII: sample split test using 25 observations for the structural equation;
- IX: sample split using 50 observations;
- X: sample split using 75 observations;
- XI: sample split using 90 observations;
- XII: two stage test ($2S$);
- XIII: test with latent variable replaced by observed vector (OLS).

9. ILLUSTRATIONS

In this section, we present empirical results on inference in two distinct economic models with latent regressors. The first example is based on Tobin's marginal q model of investment (Tobin, 1969), with fixed assets used as the instrumental variable for q. The second model stems from educational economics and relates students' academic achievements to a number of personal characteristics and other socioeconomic variables. Among the personal characteristics, we encounter a variable defined as the self esteem which we consider as latent and instrument with some measure of prestige of parents' professional occupation. The first example is one where we have good instruments, while the opposite holds for the second example.

Consider first Tobin's marginal q model of investment (Tobin, 1969). Investment of an individual firm is defined as an increasing function of the shadow value of capital, equal to the present, discounted value of expected marginal profits. In Tobin's original setup, investment behavior of all firms is similar and no difference arises from the degree of availability of external financing. In fact, investment behavior varies across firms and is determined, in great extent, by financial constraints some firms are facing in the presence of asymmetric information. For those firms, external finance may either be too costly or may not be provided for other reasons. Thus, investment depends heavily on the firm's own source of finance, namely the cash flow. To account for differences in investment behavior implied by financial constraints, several authors (Abel, 1979; Hayashi, 1985; Abel and Blanchard, 1986; Abel and Eberly, 1993; Salvas, 1995) introduce the cash flow as an additional regressor to Tobin's q model. It can be argued that another explanatory variable controlling the profitability of investment is also required. For this reason, one can argue that the firm's income has to be included in the investment regression as well. The model is thus

$$I_i = \gamma_0 + \delta Q_i + \gamma_1 CF_i + \gamma_2 R_i + e_i \tag{9.1}$$

where $\gamma = (\gamma_0,\ \gamma_1,\ \gamma_2)'$ is the coefficient vector of the matrix of independent variables $X =$ [CONST, CF, R]. The symbols CF and R denote, respectively, the cash flow and income of an individual firm. Tobin's q (denoted Q) is measured by equity plus debt and is approximated empirically by adding data on current debt, long term debt, deferred taxes and credit, minority interest and equity less inventory. Given the compound character of the variable constructed on a basis of several indexes, fixed assets are used as an explanatory variable for q in the regression which completes the model:

$$Q_i = \beta_0 + \beta_1 F_i + v_i \tag{9.2}$$

Our empirical work is based on "Stock Guide Database" containing data on companies listed at the Toronto and Montreal stock exchange markets between 1987 and 1991. The records consist of observations on economic variables describing the firms' size and performance, like fixed capital stock, income, cash flow, stock market price, etc. All data on the individual companies have previously been extracted from their annual, interim and other reports. We retained a subsample of 9285 firms which traded at the Toronto and Montreal stock exchange markets in 1991.

Since we are interested in comparing our inference methods to the widely used Wald-type tests, we first adopt the approach suggested by Pagan (1984). As the residual variance estimator obtained from the OLS regression of I on $\hat{Q}$ and the observed exogenous variables X is inconsistent,[1] Pagan proposed to use the instrumental variable or two stage least squares estimation methods. He made the point that IV and 2SLS yield correct standard errors of the parameter estimators and thus allow to obtain valid, although only asymptotically justified hypotheses test.

2SLS estimation of the model (9.1) – (9.2) means that in the first step, the dependent variable I is regressed on all the exogenous variables in the system, i.e., the constant CF, R and F while F remains the only instrument for Q. The adjusted values $\hat{Q}$ are next substituted for Q in the second stage regression. The results are summarized in Table 9.1.

[1] For a formal proof, see Pagan (1984).

Table 9.1: 2SLS Estimates, $N = 9285$

DEPENDENT VARIABLE: INVESTMENT

VAR	EST	STD. ERR	T-RATIO	P-VALUE
CONST	0.0409	0.0064	6.341	0.0000
Q TOBIN	0.0052	0.0013	3.879	0.0001
CASH FLOW	0.8576	0.0278	30.754	0.0000
INCOME	0.0002	0.0020	0.109	0.9134

DEPENDENT VARIABLE: TOBIN'S Q

VAR	EST	STD. ERR	T-RATIO	P-VALUE
CONST	1.0853	0.1418	7.650	0.0000
FIX ASST	2.4063	0.0400	60.100	0.0000

Table 9.2 below presents the parameter values computed for a subsample of 100 randomly chosen firms. These results allow us to show how the length of the confidence interval vary with the sample size.

Table 9.2: 2SLS Estimates, $N = 100$

DEPENDENT VARIABLE: INVESTMENT

VAR	EST	STD. ERR	T-RATIO	P-VALUE
CONST	0.0436	0.0376	1.158	0.2497
Q TOBIN	0.0779	0.0038	20.142	0.0000
CASH FLOW	−3.0535	0.1666	−18.325	0.0000
INCOME	0.2620	0.0309	8.471	0.0000

DEPENDENT VARIABLE: TOBIN'S Q

VAR	EST	STD. ERR	T-RATIO	P-VALUE
CONST	2.4967	2.3785	1.050	0.2964
FIX ASST	4.7814	1.4764	3.238	0.0016

The following Table 9.3 presents the 95% confidence intervals for Tobin's q. The three first intervals are obtained from, respectively, 2SLS, Two Stage and Augmented Two Stage methods by adding or subtracting 1.96 times the standard error to/from the estimated parameter value.[2] Below, we report the confidence intervals computed on the basis of the exact methods — instrument substitution and sample split, discussed in section 2. Recall that the precision of the confidence intervals depends, in the case of the sample split method, on the number of observations retained for the estimation of the structural equation. We thus show the results for, respectively, 50%, 75% and 90% of the entire number N.

Table 9.3: Confidence Intervals, $N = 9285$

METHOD	INTERVAL	
2SLS	0.0026,	0.0078
AUG. TWO STAGE	0.0025,	0.0079
TWO STAGE	− 0.0091,	− 0.0029
INSTRUMENT SUBST.	0.0025,	0.0078
SAMPLE SPLIT 50%	0.0000,	0.0073
SAMPLE SPLIT 75%	0.0017,	0.0077
SAMPLE SPLIT 90%	0.0023,	0.0078

We complete the analysis by presenting below (Table 9.4) the confidence intervals for one hundred randomly chosen firms. As it has been expected, the intervals' lengths increase with the diminishing sample size.

[2]The Augmented Two Stage method has been proposed by Pagan (1986) as an alternative for the ordinary Two Stage method to avoid underestimated standard errors. It consists here of adding the variable F, i.e., the instrument for Tobin's q, to the structural equation.

Table 9.4: Confidence Intervals, $N = 100$

METHOD	INTERVAL	
2SLS	0.0703,	0.0855
AUG. TWO STAGE	0.0618,	0.0940
TWO STAGE	− 0.1788,	− 0.1174
INSTRUMENT SUBST.	0.0693,	0.0851
SAMPLE SPLIT 50%	0.0218,	0.0191
SAMPLE SPLIT 75%	0.0697,	0.0868
SAMPLE SPLIT 90%	0.0699,	0.0860

It is clear that the confidence intervals computed on the basis of the 2SLS regression, by instrument substitution and by sample split based on 75% and on 90% of the entire sample coincide to a great extent. This result is not surprising, given the high quality of instrumental variable used for Tobin's q, as it can also be easily inferred from Tables 9.1 and 9.2.

Let us now present another example where, on the contrary, important discrepancies arise between the intervals based on the asymptotic and the exact inference methods.

Montmarquette and Mahseredjian (Montmarquette and Mahseredjian, 1989; Montmarquette, Houle, Crespo and Mahseredjian, 1989) study the student's academic achievements as a function of personal and socioeconomic explanatory variables. The student's school results in French and mathematics are measured by the grade, taking values on the interval $0 - 100$. The grade variable is assumed to depend on some student's personal characteristics, such as age, the intellectual ability (IQ) observed in the kindergarten and the self esteem measured on a adapted children self esteem scale ranging from 0 to 40. Other explanatory variables are parents' income, father's and mother's education measured in number of years of schooling, number of siblings, student's absenteeism, his own education and experience as well as the class size. We examine the significance of the self esteem variable, modelled as a latent regressor, to explain the first grader's achievements in mathematics. The self esteem of younger children has been measured, in this particular study, on a French adaptation of McDaniel–Piers scale. Noting that the measurement scale may not be equally well adjusted to the age of all students and that there exist a high degree of arbitrariness regarding the choice of the criterion, we had to slightly modify the original model: We use, namely, a set of instruments consisting of Blishen indices that reflect the prestige of father's and mother's professional occupations to get rid of eventual mismeasurement.

The data stem from a 1981–1982 enquiry of first graders attending Montreal francophone public elementary schools. The sample consists of 603 observations on students' achievements in mathematics. Given that we have to handle the limited dependent variables, we perform some necessary transformations that are explained below.

The model can be written in the following form:

$$(9.3) \quad \begin{aligned} \mathrm{LMAT} \;=\; & \beta_0 + \delta\,\mathrm{SE} + \beta_1\,\mathrm{IQ} + \beta_2\,\mathrm{I} + \beta_3\,\mathrm{FE} + \beta_4\,\mathrm{ME} + \beta_5\,\mathrm{SN} \\ & + \beta_6\,\mathrm{A} + \beta_7\,\mathrm{ABP} + \beta_8\,\mathrm{EX} + \beta_9\mathrm{ED} + \beta_{10}\,\mathrm{ABS} + \beta_{11}\,\mathrm{CS} + e_i, \end{aligned}$$

where LMAT $= \ell n(\text{grade}/(100 - \text{grade}))$, SE $= \ell n(\text{self esteem test result}/(40 -$ self esteem test result$))$, IQ is a measure of intelligence (observed in kindergarten), I is parents' income, FE and ME are father's and mother's years of schooling, SN denotes the sibling's number, A is the age of the student, ABP is a measure of teacher's absenteeism, EX indicates the years of student's work experience, ED measures his education in years, ABS is student's absenteeism and CS denotes the class size. Finally, the instrumental regression can be expressed as follows:

$$(9.4) \qquad \mathrm{SE} = \gamma_0 + \gamma_1\,\mathrm{FP} + \gamma_2\,\mathrm{MP} + v_i,$$

where FP and MP correspond to the prestige of the father and mother's profession expressed in terms of Blishen indices.

The 2SLS estimates are reported in Table 9.5 below.

Table 9.5: 2SLS Estimates, $N = 603$

DEPENDENT VARIABLE: LMAT

VAR	EST	STD. ERR	T-RATIO	P-VALUE
CONST	–4.1557	0.9959	–4.173	0.0000
SE	0.2316	0.3813	0.607	0.5438
IQ	0.0067	0.0015	4.203	0.0000
I	0.0002	0.3175	0.008	0.9939
FE	0.0015	0.0089	0.172	0.8636
ME	0.0393	0.0117	3.342	0.0009
SN	–0.0008	0.0294	–0.029	0.9767
A	0.0144	0.0070	2.050	0.0408
ABP	–0.0008	0.0005	–1.425	0.1548
EX	–0.0056	0.0039	–1.420	0.1561
ED	–0.0007	0.0206	–0.035	0.9718
ABS	–0.0001	0.0002	–0.520	0.6033
CS	–0.0184	0.0093	–1.964	0.0500

DEPENDENT VARIABLE: SE

VAR	EST	STD. ERR	T-RATIO	P-VALUE
CONST	0.8117	0.1188	6.830	0.0000
FP	0.5120	0.2625	1.951	0.0516
FM	0.6170	0.2811	2.194	0.0286

The results presented above clearly indicate that the confidence interval of the self esteem variable is finite. The exact methods, on the contrary, confirm the intuitive guess that the estimated latent variable is, in fact, far from being identified. In presence of poor instruments, we obtain infinite confidence intervals which cover the entire line of real numbers. For example, the instrument substitution method yields the confidence interval defined by the inequality

$$-31.9536\,\delta_0^2 - 84.7320\,\delta_0 - 850.9727 \leq 0$$

Since the roots of the second order polynomial are complex, and $a < 0$, the δ coefficient can take on any value between $-\infty$ and $+\infty$.

10. CONCLUSIONS

The inference methods presented in this paper are applicable to a variety of models, as regressions with unobserved explanatory variables or structural models which can be estimated by instrumental variable methods (e.g., simultaneous equations models). They may be considered as extensions of Anderson-Rubin procedures where the major improvement consists of providing tests of hypotheses on subsets or elements of the parameter vector. This is accomplished via a projection technique allowing for inference on general, possibly nonlinear transformations of the parameter vector of interest. We emphasized that our test statistics, being pivotal or at least boundedly pivotal functions, yield valid confidence sets which are unbounded with a non-zero probability. The unboundedness of confidence sets is of particular importance when the instruments are poor and the parameter of interest is hence unidentifiable or close to being unidentified. Accordingly, a valid confidence set should cover the entire set of real numbers since all values are observationally equivalent [see Dufour (1994)]. Our empirical results indicate that inference methods based on Wald-type statistics are involved in the presence of poor instruments since asymptotic tests still yield finite confidence sets. In general, unidentifiability of parameters results either from low quality instruments or, more fundamentally, from a poor model specification. A valid test yielding an infinite confidence set becomes thus a relevant indicator of problems involving the econometric setup. The power properties of exact and Wald-type tests were compared in a simulation-based experiment. The test performances were examined by simulations on a simple model with varying levels of instrument quality and the extent to which the null hypotheses differ from the true parameter value. The exact procedures were found better or, occasionally at least equally good power properties as the alternative methods.

It is important to note that although the simulations were performed under the normality assumption, our tests yield valid inferences in more general cases involving non-Gaussian errors and weakly exogenous instruments. This result has a theoretical justification and is also confirmed by our empirical examples. Since the inference methods we propose are as well computationally easy to perform, they can be considered as a reliable and a powerful alternative to Wald-type procedures.

REFERENCES

Abel, A. (1979), "Investment and the Value of Capital," New York, Garland Publishing.

Abel, A. and O.J. Blanchard (1986), "The Present Value of Profits and Cyclical Movements in Investment," *Econometrica*, 54, 249–273.

Abel, A. and J. Eberly (1993), "A Unified Model of Investment Under Uncertainty," *NBER Working Paper* 4296.

Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

Angrist, J.D. and A.B. Kruéger (1994), "Split Sample Instrumental Variables," Technical Working Paper NBER Working Paper 150, National Bureau of Economic Research, Cambridge (MA).

Barro, R.J. (1977), "Unanticipated Money Growth and Unemployment in the United States," *The American Economic Review*, Vol. 67, No. 2, 101–115.

Bound, J., D.A. Jaeger, and R. Baker (1993), "The Cure Can Be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables," Technical Working Paper 137, National Bureau of Economic Research, Cambridge (MA).

Bound, J., D.A. Jaeger, and R.M. Baker (1995), "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association* 90, 443–450.

Buse, A. (1992). "The Bias of Instrumental Variables Estimators," *Econometrica*, 60, 173–180.

Dufour, J.-M. (1994), "Some Impossibility Theorems in Econometrics with Applications to Instrumental Variables, Dynamic Models and Cointegration," *C.R.D.E. Working Paper*.

Dufour, J.-M. (1990), "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58, 475–494.

Dufour, J.-M. and J.F. Kiviet, (1994), "Exact Inference Methods in First-order Autoregressive Distributed Lag Models," Technical Report, *C.R.D.E.*, Université de Montréal, and Faculty of Economics and Econometrics, University of Amsterdam.

Fuller, W.A. (1987), "Measurement Error Models," (Wiley, New York).

Hall, A.R., Rudebusch, G.D. and D.W. Wilcox (1994), "Judging Instrument Relevance in Instrumental Variables Estimation," *Finance and Economic Discussion Series*, Division of Research and Statistics, Division of Monetary Affairs, Federal Reserve Board, Washington, D.C.

Hayashi, F. (1982), "Tobin's Marginal q and Average q: A Neoclassical Interpretation," *Econometrica*, 50, 213–224.

Hayashi, F. (1985), "Corporate Finance Side of the q Theory of Investment," *Journal of Political Economics*, 27, 261–280.

Maddala, G.S. and J. Jeong (1992), "On the Exact Small Sample Distribution of the Instrumental Variable Estimator," *Econometrica*, 60, 181–183.

Maddala, G.S. (1974), "Some Small Sample Evidence on Tests of Significance in Simultaneous Equations Models," *Econometrica*, 60, 181–183.

Murphy, K.M. and R.H. Topel (1985), "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics*, Vol. 3, No. 4, 370–379.

Montmarquette, C. and S. Mahseredjian (1989), "Could Teacher Grading Practices Account for Unexplained Variation in School Achievements?", *Economics of Education Review*, Vol. 8, No. 4, 335–343.

Montmarquette, C., Houle, R., Crespo, M. and S. Mahseredjian (1989), "Les interventions Scolaires en Milieu Defavorise: Estimation et Evaluation," Les Presses de l'Universite de Montreal.

Nagar, A.L. (1959), "The bias and Moment Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations," *Econometrica* 27, 575–595.

Nelson, C.R. and R. Startz (1990a), "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument Is a Poor One," *Journal of Business*, 63, 125–140.

Nelson, C.R. and R. Startz (1990b), "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

Oxley, L. and M. McAleer (1993), "Econometric Issues in Macroeconomic Models with Generated Regressors," *Journal of Economic Surveys*, Vol. 7, No. 1, 1–39.

Pagan, A. (1986), "Two Stage and Related Estimators and Their Applications," *Review of Economic Studies*, LIII, 517–538.

Pagan, A. (1984), "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, Vol. 25, No. 1, 221–247.

Richardson, O.H. (1968), "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, 63, 1214–1226.

Sawa, T. (1969), "The Exact Sampling Distribution of Ordinary Least Squares and Two-stage Least Squares Estimators," *Journal of the American Statistical Association*, 64, 923–937.

Staiger, D. and J.H. Stock (1993), "Instrumental Variables Regression with Weak Instruments," Technical Working Paper 151, National Bureau of Economic Research, Cambridge (MA).

Stochastic Volatility and Time Deformation: An Application to Trading Volume and Leverage Effects

1. INTRODUCTION

Asset prices respond to the arrival of information. Some days, even some parts of a trading day, very little news, good or bad, is released. Trading is typically slow and prices barely fluctuate. In contrast, when new information changes expectations, trading is brisk and the price process evolves much faster. This observation motivated Mandelbrot and Taylor (1967) and particularly Clark (1973) to suggest modelling asset price processes as subordinated stochastic processes. Instead of studying asset prices as a function of (equally spaced calendar) time, via monthly, weekly, daily or intraday series, they suggested to let asset price movements be a function of information arrival which itself evolves randomly through time. To be slightly more formal, instead of studying say daily returns as $x(\Delta t) = \log(p(t)/\, p(t-1))$, it was suggested to view $\log(p(t)/\, p(t-1) = x(T(t))$ where $T(t)$ is a positive stochastic process, sometimes called a directing process.[1] This setup, which is sometimes also called time deformation since the relevant time scale is no longer calendar time t but *operational* time $T(t)$, has several attractive features. For instance, it easily accommodates leptokurtic distributions for asset returns as emphasized by Mandelbrot and Taylor; it is also a convenient framework to study trading volume and asset return comovements as stressed by Clark; and, last but surely not least, it yields a random variance or what nowadays would be called a *stochastic volatility* model. These ideas have been refined and extended in several ways. Particularly, the restrictive assumption made in the early work that $T(t)$ was an i.i.d. process was relaxed by Tauchen and Pitts (1973). Other contributions include Harris (1987), Lamoureux and Lastrapes (1990), Gallant, Hsieh and Tauchen (1991), Andersen (1993).[2] Moreover, the microstructure foundations for time deformation and the process of price adjustments can be found most explicitly in Easley and O'Hara (1992). It is interesting and at the same time important to note that none of these developments exploited explicitly the continuous time financial modelling approach which has become so widely used since the seminal work of Merton (1973) and many others. Indeed, when one refers to *stochas-*

[1] Clark, for instance, deliberately chose the notation $T(t)$ to indicate he meant the trading volume on day t.

[2] There is, of course, also an extensive literature on trading volume, including both theoretical and empirical papers. See, for instance, Foster and Viswanathan (1993a,b), Gallant, Rossi and Tauchen (1992), Hausman and Lo (1991), Huffman (1987), Karpoff (1987), Lamoureux and Lastrapes (1993), Wang (1993), among others.

tic volatility, one typically thinks of models originally constructed for the valuation of options where changes in the volatility are governed by a stochastic differential equation which is not explicitly related to the arrival of information through trading volume or other variables. Such models were developed by Hull and White (1987), Johnson and Shanno (1987), Scott (1987), Wiggins (1987), Chesney and Scott (1989), Stein and Stein (1991) and Heston (1993), among others.

In this paper, we study continuous time stochastic volatility models with time deformation. The setup combines insights borrowed from the earlier literature on subordinated stochastic processes and from the more recently developed diffusion equation stochastic volatility models. Let us briefly return to a more formal discussion and note that the latter class of models typically takes the form:

(1.1a) $$dy(t) = \mu y(t)dt + \sigma(t)y(t)dW_1(t)$$

(1.1b) $$d\log\sigma(t) = a(b - \log\sigma(t))dt + cdW_2(t)$$

where $W_1(t)$ and $W_2(t)$ are two standard Wiener processes usually assumed to be independent. We will not assume that the volatility process moves continuously and smoothly through calendar time, as is usually assumed and described by (1.1b). The initial motivation for the work of Mandelbrot and Taylor, as well as Clark, was that key variables affecting volatility, like the arrival of information to the market, tend not to evolve continuously and smoothly through time. Therefore, we shall make the volatility process a subordinated stochastic process evolving in a time dimension set by market activity. To make this more explicit, let us assume an operational time scale s for the volatility process, with $s = g(t)$, a mapping between operational and calendar time t, such that:[3]

(1.2a) $$dy(t) = \mu y(t)dt + \sigma(g(t))y(t)d\omega_1(t)$$

(1.2b) $$d\log\sigma(s) = a(b - \log\sigma(s))ds + cd\omega_2(s)\,.$$

[3]The mapping $s = g(t)$ must satisfy certain regularity conditions which will be discussed later.

Following Stock (1988), we denote by $g(t)$ the directing process. It may depend on trading volume besides many other series that help determine the pace of the market. Before discussing this issue in detail, we would like to make some observations regarding equations (1.2). Indeed, it should first be noted that the equations collapse to the usual stochastic volatility model if $g(t) = t$. This was done on purpose to accommodate econometric hypothesis testing. Obviously, we could have adopted another specification for $\sigma(g(t))$. Moreover, one could correctly argue that, defining volatility as a subordinated process amounts to suggesting a more complex law of motion for volatility in comparison to the Ornstein-Uhlenbeck (henceforth $O - U$) specification appearing in (1.1b). This interpretation is valid, yet it should be noted that, through $g(t)$, one can associate many series other than the security price $y(t)$ to explain volatility; hence, one implicitly deals with a multivariate framework.

What should determine $g(t)$? Since the flow of information is a latent process, we have to specify the mapping $s = g(t)$ in terms of related observable processes. We propose to use past volume of trade and other variables such as past returns allowing possibly for an asymmetric response to create a leverage effect. Therefore, our setup provides a way of introducing data on trading volume in the specification of stochastic volatility models. Furthermore, it is possible to accommodate leverage effects by specifying a directing process that would allow for asymmetric responses of s to past price changes. Hence, out operational time evolves differently in bull and bear markets. The empirical results suggest that our specification provides an alternative to a class of option pricing models, put forward by Merton (1976a, b), where jumps in the underlying security returns are permitted. Merton suggested to include a Poisson jump process to distinguish between the arrival of *normal* information, modeled as a standard log normal diffusion, and the arrival of *abnormal* information, modeled as a Poisson process. We find that operational time typically moves slowly, but every so often one finds dramatic increases in market speed. In Merton's setup, the information arrival spells are purely exogenous, whereas our approach has the sources of these changes modeled both in a multivariate sense, via the introduction of volume series, and in an endogenous fashion through past price changes. Using daily $S\&P$ 500 data and NYSE volume from 1950–1987, we find that increases in volume accelerate operational time, resulting in volatility being higher, less persistent and subject to shocks with a higher innovation variance. Downward price movements have similar effects, while upward price movements increase persistence in volatility and decrease the dispersion of shocks by slowing down the operational time clock.[4]

In order to estimate the subordinated diffusions, we rely on two alternative estimation procedures. The first method involves the Kalman filter and draws upon results from Harvey, Ruiz and Shephard (1994) on estimating stochastic volatility models in state space form and results from Stock (1988) on estimation of linear processes with time deformation. The second method is based on a "matching moment" principle as presented by Gallant and Tauchen (1994) using SNP densities for stock returns and volume series fitted by Gallant, Rossi and Tauchen (1992).

In section 2, we present the basic model. Estimation and hypothesis testing are discussed in section 3. Empirical results appear in section 4.

2. A TIME DEFORMATION APPROACH TO STOCHASTIC VOLATILITY

Stochastic processes used in finance are most often assumed to be generated by a first-order stochastic differential equation of the form:

$$dX(s) = a(s,\, X(s),\, \ominus)ds + b(s,\, X(s),\, \ominus)dM(s) \tag{2.1}$$

where $X(s)$ is a n-dimensional process adapted to a filtered probability space (Ω, F, P) evolving in some *operational* time. The process is parameterized by $\ominus \in \mathbf{R}^P$ with $dM(s)$ a m-dimensional semi-martingale process, while $a(s, X(s), \ominus)$ and $b(s, X(s), \ominus)$ are both bounded predictable processes of dimensions n and $n \times m$, respectively. Equations like (2.1) have been adopted to describe security, bond and derivative prices as well as information flows, mortgage values, inventories and other state variables such as technology. Whenever the assumed operational time scale s differs from t, there is so-called time deformation or, alternatively with $s = g(t)$, the process $X(g(t))$ is a subordinated stochastic process. Both expressions will be used throughout the paper.

[4] Obviously other series could figure in the specification of $g(t)$. Indeed in many instances one can find time deformation arguments in financial modelling. In section 2, we will provide a brief review of examples which appeared in the literature.

Besides the aforementioned works of Clark, Mandelbrot and Taylor, the idea of time deformation appears in quite a number of finance papers, though not always explicitly. Probably the simplest examples of time deformation are related to the widely documented nontrading day, holiday and weekend effects in asset prices. Bessembinder and Hertzel (1993) are the most recent example of several papers on these so-called stock market anomalies.[5] In foreign exchange markets, *activity* scale determined by the number of active markets around the world at any particular moment can be considered. Dacorogna et al. (1993) describe explicitly a model of time deformation along these lines for intraday movements of foreign exchange rates. Besides these relatively simple examples, there are a number of more complex ones. The most prominent being the work of Mandelbrot and Taylor as well as Clark and extensions which were mentioned in the introduction. Before elaborating on this further, it is worth mentioning a few other examples as well. For instance, Madan and Seneta (1990) and Madan and Milne (1991) introduced a Brownian motion evaluated at random (exogenous) time changes governed by independent gamma increments as an alternative martingale process for the uncertainty driving stock market returns. Geman and Yor (1993) also used time-changed Bessel processes to compute path-dependent option prices such as is the case with Asian options.[6]

As explained in the introduction, we study a continuous time stochastic volatility model with time deformation. We combine the insights from Mandelbrot and Taylor (1967) and Clark (1973) on subordinated stochastic processes and from the option pricing stochastic volatility models associated with the work of Hull and White (1987) and others mentioned before. Volatility is modeled as a subordinated process driven by a generic directing process $s = g(t)$, s being an operational time scale, associated with the arrival of information. In particular, we consider the set of equations (1.2) repeated here for convenience:

[5]For instance, Lakonishok and Smidt (1988) and Schwert (1990) argue that returns on Monday are systematically lower than any other day of the week, while French and Roll (1986), French, Schwert and Stambaugh (1987) and Nelson (1991) demonstrate that daily return volatility on the NYSE is higher following nontrading days closures.

[6]Time deformation is also used for a variety technical reasons in, for instance, Detemple and Murthy (1993) to characterize intertemporal asset pricing equilibria with heterogeneous beliefs. Nelson and Foster (1993, 1994) use changes in time scales to study ARCH models as filters for diffusion models.

$$(2.2a) \qquad dy(t) = \mu y(t)dt + \sigma(g(t))y(t)dW_1(t)$$

$$(2.2b) \qquad d\log\sigma(s) = a(b - \log\sigma(s))ds + cdW_2(s)\,.$$

To enhance our understanding of the mechanics of the process, let us momentarily isolate the volatility equation (2.2b) and discuss its properties as well as its discretization. To simplify this task even further, let us set $b = 0$ and work with a continuous time AR(1). To describe an investor's information, let us consider the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and the nondecreasing family $F = \{\mathcal{F}_t\}_{t=0}^{+\infty}$ of $sub - \sigma$-algebras in calendar time. Furthermore, let Z_t be a m-dimensional vector process adapted to the filtration F, i.e., Z_t is $\mathcal{F}_t$-measurable. The rate of operational time speed will be assumed to be $\mathcal{F}_{t-1}$ measurable via the logistic transformation:

$$(2.3) \qquad \frac{dg(\tau; Z_{t-1})}{d\tau} \equiv \dot{g}(\tau,\, Z_{t-1}) \equiv \exp(c'Z_{t-1}) \,/\, \left\{\frac{1}{T}\sum_{t-1}^{T}\exp(c'Z_{t-1})\right\}$$

for $t-1 \leq \tau < t$.[7] Equation (2.3), setting the speed of change of operational time as a measurable function of calendar time process Z_{t-1}, is complemented with additional identification assumptions:

$$(2.4a) \qquad 0 < \dot{g}(\tau;\; Z_{t-1}) < \infty$$

$$(2.4b) \qquad g(0) = 0$$

$$(2.4c) \qquad \frac{1}{T}\sum_{t=1}^{T}\Delta g(t) = 1\,.$$

[7]The fact that the denominator in (2.3) contains a sample average may suggest that $\sigma(g(t))$ is not measurable with respect to the filtration $\mathcal{F}_t$ in calendar time. However, the denominator in (2.3) is there for reasons of numerical stability of the algorithms described in the next section. Since it is only a scaling factor, its presence is of no conceptual importance.

These three technical conditions, which will not be discussed at length here as they are covered in detail in Stock (1988), guarantee that the operational time clock progresses in the same direction as calendar time without stops or jumps.[8] Given that $\dot{g}$ is constant between successive calendar time observations via (2.3), its discrete time analogue $\Delta g(t) \equiv g(t) - g(t-1)$ takes the same logistic form appearing in (2.3). At this point, we have not yet discussed what series should enter the vector Z_{t-1}. A detailed discussion will be delayed until later in the section, but it may be worth pointing out at those variables like past trading volume or any other processes linked to information arrival will be candidate series to enter equation (2.3). Proceeding with the discussion of equation (2.2b), we note that the solution in operational time of a first-order linear process can be expressed as:

(2.5) $$\log \sigma(s) = e^{a(s-s')} \log \sigma(s') + \int_{s'}^{s} e^{a(s-r)} dW_2(r)$$

where $s' < s$. To recover the solution in calendar time, we let $s = g(t)$ and $s' = g(t-1)$ and obtain:

(2.6a) $$h_t = e^{a\Delta g(t)} h_{t-1} + \nu_t \qquad t = 1, \ldots, T$$

(2.6b) $$\nu_t \sim N(0, -\sum (1 - \exp(2a\,\Delta g(t)))\,/\,2a)$$

(2.6c) $$\Delta g(t) = \exp(c' Z_{t-1})\,/\,\frac{1}{T}\left\{\sum_{t=1}^{T} \exp(c' Z_{t-1})\right\}$$

Hence, the process where $h_t = log\sigma(g(t))$ while linear in operational time becomes a random coefficient model, also called doubly stochastic process, in calendar time also featuring conditional heteroscedasticity governed by $\Delta g(t)$.[9]

[8] Excluding jumps for the time deformation process must not be confused with the presence of jumps in the stock return process, as proposed by Merton (1976a, b). The time deformation will govern the (stochastic) volatility of the return process. Arbitrarily large (yet finite) changes in operational time will make the stock return process extremely volatile through the conditional variance.

[9] Doubly stochastic processes have been discussed in detail by Tjøstheim (1986). Stability conditions and existence of moments have been studied for cases where $\Delta g(t)$ is Markovian. It may be worth noting at this point that the Z_{t-1} process need not be exogenous. Indeed, Stock (1988) showed that by setting Z_{t-1} equal to the square of the process appearing in the mean equation, one obtains an ARCH-like process having the additional feature of a random coefficient model.

The brief digression on time deformation facilitates the presentation of the process of main interest, which is a SV model with time deformation. Suppose now that $\{y_t\}$ represents a discrete time sample of the process in (2.2). A standard Euler approximation to (2.2a) yields:

$$\text{(2.7)} \qquad \log y_t = \lambda + \log y_{t-1} + \sigma_t \varepsilon_t \qquad \varepsilon_t \sim \text{i.i.d. } N(0,\, 1)\,.$$

Let us combine this expression with the volatility equation. If we assume again that $\beta \neq 0$ in (2.2b) and furthermore observe that the stock returns used in our empirical application exhibit small yet significant autocorrelation at lag 1, we obtain the following discrete time representation:

$$\text{(2.8a)} \qquad \Delta \log y_t - a_1 \,\Delta \log y_{t-1} - \lambda = \sigma_t \,\varepsilon_t$$

$$\text{(2.8b)} \qquad h_t = (1 - \exp(a\Delta g(t))b + \exp(a\Delta g(t))h_{t-1} + \nu_t$$

where the variance of ν_t is given by (2.6b). Equations (2.8a) and (2.8b) are the basic set of equations of the discrete time representations of the SV model with a subordinated volatility process which evolves at a speed set by $\Delta g(t)$. The set of equations (2.8) will be used for a simulated method of moments estimation procedure which will be discussed in the next section. We will also use a quasi-maximum likelihood estimation algorithm, however, based on a Kalman filter state space representation. For this, we rely on Harvey, Ruiz and Shephard (1994) and write equations (2.8a) as:

$$\text{(2.9)} \qquad \log[\Delta \log y_t - a_1 \Delta \log y_{t-1} - \lambda]^2 = h_t + \log \varepsilon_t^2$$

where $E \log \varepsilon_t^2 = -1.27$ and $\text{Var} \log \varepsilon_t^2 = \pi^2/2$. We can rewrite equation (2.9) adding (2.8b), as follows:

$$\text{(2.10a)} \qquad \log[\Delta\, log\, y_t - a_1 \,\Delta \log y_{t-1} - \lambda]^2 = -1.27 + h_t + \zeta_t$$

(2.10b) $$h_t = (1 - \exp(a\Delta g(t))b + \exp(a\Delta g(t))h_{t-1} + \nu_t\,.$$

Except for the parameter λ, whose treatment is discussed, for instance, by Gouriéroux, Monfort and Renault (1993), we obtain a state-space model with time-varying coefficients similar to that obtained by Stock (1988), apart from the properties of the ζ_t process which is no longer Gaussian.[10] Consequently, the estimation procedure based on the Kalman filter will result here in a quasi-maximum likelihood estimator, similar to Harvey, Ruiz and Shephard (1994).

Obviously, the SV model with time deformation can be viewed simply as a model with a doubly stochastic process for h_t which replaces the usual linear or Ornstein-Uhlenbeck process. Yet, the stochastic variation in the autoregressive coefficient has a very specific interpretation through the specification of the mapping $g(t)$. Let us, therefore, turn our attention now to a description of the functional form that will be adopted. Z_{t-1} should include both past volume and price movements. With respect to price movements, we will adopt a functional form which can allow for asymmetries in the time deformation when prices move upward or downward. Finally, we could also include a set of predetermined processes denoted d_t to account for nontrading day effects and possibly other periodic patterns discussed, for instance, by Bollerslev and Ghysels (1994). The logistic function emerging from the above discussion would be:[11]

(2.11) $$\exp(c'Z_{t-1}) \equiv \exp(c_d'\, d_t + c_v \log \mathrm{Vol}_{t-1} + c_p\,\Delta \log y_{t-1} + c_\ell\,|\,\Delta \log y_{t-1}\,|)\,.$$

where $\log \mathrm{Vol}_t$ represents a trading volume series and $\Delta \log y_t$ the return series. The specification of the time deformation function is chosen in light of certain existing stylized facts we would like the model to fit. Other specifications can be chosen, however. The general model we develop holds for any process Z_{t-1}, which is assumed

[10] The innovations ν_t and ζ_t are assumed *i.i.d.* Correlation between the two processes would create asymmetries in the conditional variance [see Harvey and Shephard (1993)]. We do not need to assume such a correlation since the asymmetry will come through the time deformation (as will be discussed later in the text).

[11] Note that the timing of d_t differs from the other processes. Since the variables entering d_t are predetermined, they are measurable with respect to $\mathcal{F}_{t-1}$ and, therefore, legitimate for setting the pace of operational time changes $\Delta g(t)$. Moreover, it should be observed that c_d is a vector of parameters since d_t may be multivariate.

to capture the flow of information. The specification in (2.11) is just one of possibly many, yet is directly related to the existing literature on conditional variance models. Further research may find other series appropriate as well.

It might be useful to describe the stochastic behavior of the process obtained so far. Referring to some of the empirical results, discussed later, we must first observe that coefficient A in (2.8b) is found to be negative. Therefore, when $c_v > 0$, the model predicts that increases in volume make $\Delta g(t)$ increase. This acceleration in operational time results in a decline in $a_t \equiv \exp A\Delta g(t)$ and an increase in $\sigma^2_{\nu t}$ defined in (2.6b). These two effects imply that the h_t process becomes more erratic since its persistence declines and it is subject to larger shocks. Thus, trading volume increases are paired with volatility increases, an empirical fact documented via SNP fitting by Gallant, Rossi and Tauchen. If we find $c_p < 0$ combined with $c_\ell > 0$, while $|c_\ell| > |c_p|$ to ensure $\Delta g(t) > 0$, then a change in price of the same magnitude but of the opposite sign will result in $\Delta g(t)$ to be smaller with upward price movements and larger with falling prices. Consequently, declining stock prices have an effect of making the volatility process more erratic (i.e., a_t declines and $\sigma^2_{\nu t}$ increases), while a positive price move of the same size has an opposite effect, namely, a_t increases and $\sigma^2_{\nu t}$ decreases.

When the subordinated stochastic volatility model involves trading volume through $\Delta g(t)$ it is natural to consider a bivariate model of stock returns and trading volume since both series are jointly determined by the arrival of information. It is indeed a major point stressed by Clark (1973), Tauchen and Pitts (1983) and many others. The framework developed so far lends itself easily to extensions which take into account the laws of motion of trading volume. Such model would be as follows:

$$(2.12a) \qquad \begin{pmatrix} (\Delta \log y_t - a_1 \Delta \log y_{t-1}) \\ \log \mathrm{Vol}_t \end{pmatrix} = \begin{pmatrix} \mu_P \\ \mu_V \end{pmatrix} + \begin{pmatrix} \sigma_t \varepsilon_t \\ v_t \end{pmatrix}$$

$$(2.12b) \qquad \begin{pmatrix} h_t \\ v_t \end{pmatrix} = [I - \exp(A\Delta g(t))] \begin{pmatrix} \beta_P \\ \beta_V \end{pmatrix} + \exp(A\Delta g(t)) \begin{pmatrix} h_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix}$$

where:

$$(2.13) \qquad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

$$E\nu_t\nu_t' = \Sigma[(I - \exp(A\Delta g(t))]A^{-1/2} \tag{2.14}$$

$$\sum = \begin{pmatrix} \sigma_P^2 & \rho \\ \rho & \sigma_\tau^2 \end{pmatrix} \tag{2.15}$$

Equation (2.14) is the bivariate extension of (2.6b) now involving the matrix A defined in (2.13) and the covariance matrix Σ. Note also that the time deformation $\Delta g(t)$ is common to both processes. It is clear that while the specification of $\Delta g(t)$ remains the same, the arrival of information which drives jointly h_t and v_t is different in comparison to the univariate model only involving h_t. It should also be noted the past volume affects h_t directly through the term $a_{12}\Delta g(t)v_{t-1}$ as well as through $\Delta g(t)$. The dynamics of the joint process (2.12) will be more difficult to describe. Instead, our discussion will revolve around the analysis of impulse response functions for the nonlinear bivariate system along the lines suggested by Gallant, Rossi and Tauchen (1993) and Potter (1991).

3. ECONOMETRIC ANALYSIS

Estimating SV models represents some stiff challenges for econometricians. In recent years, several estimation principles were proposed involving the use of simulated method of moments, Kalman filter and Bayesian procedures. Recent contributions include Duffie and Singleton (1993), Gallant and Tauchen (1994), Gouriéroux, Monfort and Renault (1993), for the method of moments estimators while Harvey, Ruiz and Shephard (1994) and Jacquier, Polson and Rossi (1994) discuss, respectively, the Kalman filter and Bayesian methods. To estimate the SV models with time deformation, we shall adopt two methods: one using the Kalman filter and one relying on the "matching moments" approach described by Gallant and Tauchen (1994). A subsection will be devoted to each method. Before turning to the specifics, it is worth making several observations. Both estimation procedures should be viewed as complementary especially with regard to estimating subordinated processes. The Kalman filter estimator is a quasi-maximum likelihood procedure, henceforth QMLE, and therefore has the disadvantage of being asymptotically inefficient. Simulation evidence reported in Andersen and Sorensen (1994) and Jacquier, Polson and Rossi (1994) suggests that the state space QMLE may be quite inefficient, depending on the circumstances. This setup has certain advantages, however, in comparison to the simulated methods of moments procedure. Indeed, there is a greater flexibility with the Kalman filter in formulating $\Delta g(t)$ without having to match the moments of all the series involved in the time scale transformation. Hence, there are certain trade-offs between the two estimation procedures which we will discuss. Therefore, we turn our attention now to the specific details to clarify these observations.

3.1 Quasi-Maximum Likelihood Estimation of SV Models with Time Deformation

This method consists of maximizing the quasi-likelihood function of a nonlinear SV model written in a form of linear discrete-time state space system as specified in equations (2.10a and b). The Gaussian quasi-likelihood function is evaluated in the (calendar) time domain using a Kalman filter with time varying filter parameters that depend on $\Delta g(t)$. This algorithm is described in Stock (1988) and summarized in this section. We cast the presentation in a general multivariate context since we also want to cover the bivariate model involving volume described by equation (2.12).

The evolution of the state is described by the transition equation. In operational time s, the r^{th}-order linear differential equation representing a n-dimensional $O-U$ process can be written in a stacked form as:

$$d\psi^*(s) = A[R\beta - \psi^*(s)]ds + Rd\eta(s)\,, \tag{3.1.1}$$

where

$$\psi^*(s) = \begin{bmatrix} \xi(s) \\ D\xi(s) \\ \vdots \\ D^{r-1}\xi(s) \end{bmatrix}, \quad R = \begin{bmatrix} 0 \\ \vdots \\ I \end{bmatrix}, \quad A = \begin{bmatrix} 0 & & & \\ \vdots & & I & \\ 0 & & & \\ A_r & A_{r-1} & \ldots & A_1 \end{bmatrix}.$$

Here $\xi(s)$ may represent any subordinated process of interest while the vector $\psi^*(s)$ is of dimension nr and the matrix R is $nr \times n$. The matrix of coefficients A is of dimension $nr \times nr$, its elements being $n \times n$, while the mean vector β is $n \times 1$. We denote the mean-square differential operator by D. The innovation process $\eta(s)$ is Gaussian with zero-mean increments and covariance matrix $E[d\eta(s)d\eta(s')'] = \sum ds$ for $s = s'$ and 0 otherwise. The real parts of the roots of matrix A are required to be negative for stability. We will also assume that they are distinct in order to adopt a useful eigenvalue decomposition $A = G\Lambda G^{-1}$, where Λ is a diagonal matrix of eigenvalues of A, which are, in general, complex numbers, while G is a matrix of eigenvectors of A. Following Stock (1988), we set $\psi(s) = G^{-1}\psi^{*}(s)$ and observe that in operational time the transformed variable satisfies the following equation:

$$\psi(s) = \left[I - e^{\Lambda(s-s')}\right] G^{-1}R\beta + e^{\Lambda(s-s')}\psi(s') + \int_{r=s'}^{s} e^{\Lambda(s-r)}G^{-1}Rd\eta(r)\,, \tag{3.1.2}$$

where $s > s'$. Let the calendar time state vector be denoted $S(\tau) = \psi(g(\tau))$. Evaluating the previous equation at $s = g(\tau)$ and $s' = g(t-1)$, we find that $S(\tau)$ satisfies:

$$S(\tau) = \left[I - e^{\Lambda(g(\tau)-g(t-1))}\right] G^{-1}R\beta + e^{\Lambda(g(\tau)-g(t-1))}S(t-1) + \int_{r=g(t-1)}^{g(\tau)} e^{\Lambda(g(\tau)-r)}G^{-1}Rd\eta(r)\,. \tag{3.1.3}$$

Developing the first term on the $r.h.s$ of (3.1.3), we obtain:

$$S(\tau) = G^{-1}R\beta - e^{\Lambda(g(\tau)-g(t-1))}G^{-1}R\beta + e^{\Lambda(g(\tau)-g(t-1))}S(t-1) + \int_{r=g(t-1)}^{g(\tau)} e^{\Lambda(g(\tau)-r)}G^{-1}Rd\eta(r)\,, \tag{3.1.4}$$

and hence,

$$S(\tau) - G^{-1}R\beta = e^{\Lambda(g(\tau)-g(t-1))}[S(t-1) - G^{-1}R\beta] + \int_{r=g(t-1)}^{g(\tau)} e^{\Lambda(g(\tau)-r)}G^{-1}Rd\eta(r)\,. \tag{3.1.5}$$

Now, set $\tilde{S}(\tau) = S(\tau) - G^{-1}R\beta$. It is easy to note that equation (3.1.5) can be written as:

$$\tilde{S}(\tau) = e^{\Lambda(g(\tau)-g(t-1))}\tilde{S}(t-1) + \int_{r=g(t-1)}^{g(\tau)} e^{\Lambda(g(\tau)-r)}G^{-1}Rd\eta(r)\,. \tag{3.1.6}$$

Equation (3.1.6) evaluated at $\tau = t$ yields the final representation of the transition equation:

$$\tilde{S}_t = T_t\,\tilde{S}_{t-1} + \nu_t\,, \tag{3.1.7}$$

where $T_t = \exp(\Lambda\,\Delta g(t))$ and $\nu_t = \int_{r=g(t-1)}^{g(\tau)} \exp[\Lambda(g(t)-r)]G^{-1}Rd\eta(r)$.

The multivariate measurement equation can be written, in terms of the state vector $\tilde{S}_t$, as:

$$Y_t = -1.27\imath + G\tilde{S}_t + R\beta + \zeta_t\,, \tag{3.1.8}$$

where Y_t and ζ_t are $n \times 1$ vectors with elements $y_{it} = \log[\Delta \log y_{it} - \lambda]^2$, $\zeta_{it} = \log \varepsilon_{it}^2 + 1.27$, $i = 1, \ldots, n$, and $\imath$ is a n vector of ones. The equations (3.1.7) and (3.1.8) form a linear state space system. We suppose that the disturbances in both equations are uncorrelated since in our setup any eventual impact of prices on volatility is channelled through the time deformation term $\Delta g(t)$.

The next step of the procedure consists of applying the Kalman filter algorithm. Note, however, that ζ_{it} in (3.1.8) are not normally distributed and, hence the linear filtering method can only be approximate while estimation will be asymptotically inefficient (see Andersen and Sorensen (1994) and Jacquier et al. (1994) for further discussion of this issue as well as simulation evidence).

Following Stock (1988), we initialize the Kalman filter by taking unconditional expectations and assuming that prior to the sample $\Delta g(t) = 1$. The one-step ahead forecast of the state, $a_{1|0}$ is equal to zero and its covariance matrix $P_{1|0} = \sum_{i=0}^{\infty} T^i Q \overline{T}'^i$ can be easily obtained by computing $T = T_t$, and $Q = Q_t = E(\nu_t\,\overline{\nu}_t')$ evaluated at $\Delta g(t) = 1$. Moreover, the $(i-j)$ the element of the matrix Q_t is known to be equal to $q_{ij}\int_{r=0}^{\Delta g(t)} \exp[(\lambda_i + \overline{\lambda}_j)\,(\Delta g(t) - r)]dr = -q_{ij}(1 - T_{it}\overline{T}_{jt})/(\lambda_i + \overline{\lambda}_j)$, where q_{ij} is the $(i-j)$ element of the matrix $G^{-1}R \sum R'\overline{G}^{-1'}$.

3.2 Simulated Method of Moments Estimation of Subordinated Stochastic Processes

Gallant, Rossi and Tauchen (1992) have analyzed stock returns and volume of transactions data and estimated semi-nonparametric densities, henceforth SNP, of the joint process. These SNP densities will be the setting for the simulated method of moments procedure described in this section. The moment matching procedure involving S-NP densities, dubbed by Gallant and Tauchen (1994) as efficient GMM (henceforth EMM), allows one to avoid problems related to the appropriate choice of moments in a standard GMM setup. The choice of moments is indeed particularly cumbersome in cases of highly nonlinear models, such as SV models. The EMM procedure relies on moment conditions which are generated in a first step using the score function of the auxiliary SNP model. To facilitate the presentation, let us denote the parameter vector describing the SNP density as θ while the vector α describes the parameters of the SV model. For the EMM method, we are interested in generating the vector of moment conditions using expectations under the SV model of the score from an auxiliary SNP model. The task of computing this expectations vector is facilitated since it is obtained easily by simulating the realizations, for a given value of the parameter vector α, of the SV model with deformation of time. To be more formal, consider the mapping obtained through simulation:

(3.2.1)
$$\alpha \rightarrow \{\hat{y}_\tau(\alpha),\ \hat{x}_{\tau-1}(\alpha)\}_{\tau=1}^{N} \ ,$$

where $\hat{y}_\tau$ and $\hat{x}_\tau$ denote respectively the set of simulated endogenous variables and the set of lagged endogenous variables both generated by the time deformation SV model. It is worth pointing out that we no longer rely on the normal approximation, as appearing in (2.10a), but instead use directly (2.8a). In our case, $\hat{y}_\tau(\alpha)$ would typically contain stock returns and trading volume while $\hat{x}_\tau(\alpha)$ consists of their past realisations. The estimation is performed in two steps. First, the estimation of the auxiliary model (called the score generator) yields:

(3.2.2)
$$\tilde{\Theta}_n = \underset{\Theta\, \epsilon\, \theta}{\text{Argmax}}\ \frac{1}{n} \sum_{t=1}^{n} \ell n\, f_t(\tilde{y}_t \,|\, \tilde{x}_{t-1},\, \Theta)\,,$$

where $\{\tilde{y}_t,\ \tilde{x}_{5-1}\}_{t=1}^{n}$ denotes the set of observed data from a sample of size n (the simulated data set is of size N). In the second step, the following moment criterion is computed:

$$(3.2.3) \qquad m_n(\alpha, \tilde{\ominus}_n) = \frac{1}{n}\frac{1}{N}\sum_{\tau=1}^{n} \frac{\partial}{\partial_\ominus} \; \ell n \; f_t(\hat{y}_\tau(\alpha)|\hat{x}_{\tau-1}(\alpha), \tilde{\ominus}_n) \, .$$

Finally, the estimation of the time deformation SV model is given as:

$$(3.2.4) \qquad \hat{\alpha}_n = \underset{\alpha}{\text{Argmin}} \; m_n'(\alpha, \tilde{\ominus}_n)\,(\tilde{I}_n)^{-1} m_n(\alpha, \tilde{\ominus}_n) \, ,$$

where $\tilde{I}_n$ is a weighting matrix. [For a discussion of the appropriate choice of the $\tilde{I}_n$ estimator, see Gallant and Tauchen (1994)]. The efficiency of EMM depends, of course, on the choice of the auxiliary model. A score generator nesting the time deformation SV model would allow to attain the maximum likelihood efficiency. However, even a score generator that only closely approximates the actual distribution of the data is nearly fully efficient. As noted before, the score function selected for the time deformation SV model is the derivative of the log density estimated by a semi-nonparametric method (SNP) proposed by Gallant, Rossi and Tauchen (1992). The SNP density function is based on a hermite expansion of the form:

$$(3.2.5) \qquad h(z) \propto [P(z)]^2 \Phi(z) \, ,$$

where z denoted a M-dimensional vector, $P(z)$ is a multivariate polynomial of degree K_z and $\Phi(z)$ denotes the density function of the multivariate Gaussian distribution with mean zero and an identity covariance matrix. The constant of proportionality $1 / \int [P(s)]^2 \Phi(s) ds$ makes $h(z)$ integrate to one. A more complex specification can easily be handled by means of a change of variables $y = Rz + \mu$, where R is defined as an upper triangular matrix and μ is a vector of dimension M. In consequence, we obtain the following expression:

$$(3.2.6) \qquad f(y \mid \ominus) \propto \{P[R^{-1}(y-\mu)]\}^2 \{\Phi[R^{-1}(y-\mu)] / \mid \det(R) \mid\} \, ,$$

where the leading term is now proportional to the multivariate Gaussian density function with mean μ and covariance matrix $\Sigma = RR'$. Hence, by setting K_z equal to zero, a multivariate normal density can be estimated. Nonzero values of K_z result in shape modifications that can accommodate fat tails and skewness. Other modifications can be done by adjusting values of the remaining turning parameters L_μ, L_p, L_r, K_x, I_z and I_x. The tuning parameter L_μ determines the number of lags in the location shift

μ considered as a linear function of L past values of y for accommodating a Gaussian VAR specification. To approximate a conditionally heterogeneous process, each coefficient of the polynomial $P(z)$ can be defined as a polynomial of degree K_x in past values of y. The new polynomial $P(z,\ x)$, where x is the vector of lagged values of y is hence of degree $K_z + K_x$. The conditional heteroscedasticity can also be captured by letting R be a linear function of past values of y. The number of lags in the scale shift R_x is determined by the parameter L_r. The number of lags in the x part of the polynomial $P(z,\ x)$ is controlled by L_p. Finally, the parameters I_z and I_x allow to suppress an excessive number of cross product terms in case of multivariate series.

The optimal score selection strategy is summarized in Table 3.1. The data used consist of the daily closing value of the $S\&P$ composite stock index and the daily volume of shares traded on the NYSE. The data set is identical to that used by Gallant, Rossi and Tauchen (1992), who described its sources in detail.[12] To determine the best fit, we computed three criteria: AIC, BIC and HQ. AIC is the Akaike criterion defined as:

$$\text{AIC} = S_n(\hat{\Theta}) + P_\Theta \,/\, n\,,$$

where $S_n(\hat{\Theta}_n)$ is the mean of the log likelihood and P_Θ denotes the number of parameters. The more conservative criterion of Schwartz (BIC), which penalizes specifications involving too many parameters is computed as:

$$\text{BIC} = S_N\,(\hat{\Theta}) + \tfrac{1}{2}\,(P_\Theta \,/\, N)\,\log\,(N)\,.$$

Finally, the Hannan-Quinn criterion lies in between the last two and is given by the following expression:

$$\text{HQ} = S_n(\hat{\Theta}) + (P_\Theta \,/\, n)\log\,[\log(n)]\,.$$

We retained the best performing specification under the most conservative Schwartz criterion. Our preferred SNP model is described by the following set of the tuning parameter values: $\{L_\mu = 2,\ L_r = 16,\ L_p = 2,\ K_z = 4,\ K_x = 1,\ I_z = I_x = 0\}$ with total number of 35 parameters.

Before turning to the empirical results, it may be worth pointing out that the fitted SNP densities reported in Table 3.1 only involve stock returns. A consequence of this

[12]Besides the description of data sources, they also describe several transformations.

is that the matching moment SV estimates will be limited to a time deformation based on past returns, since trading volume does not figure in the SNP density. We

Table 3.1

Univariate Price Changes 1950–1987 Optimized Likelihood

L_μ	L_r	L_p	K_z	I_z	K_x	I_x	S_N	P_Θ	BIC	HQ	AIC
2	16	2	4	0	0	0	1.2849	25	1.2963	1.2904	1.2874
2	16	2	4	0	1	0	1.2793	35	1.2955	1.2871	1.2828
2	16	4	4	0	0	0	1.2849	25	1.2963	1.2904	1.2874
2	16	4	4	0	1	0	1.2759	45	1.2969	1.2861	1.2805
2	16	6	4	0	1	0	1.2746	55	1.3003	1.2871	1.2802
2	16	8	4	0	1	0	1.2735	65	1.3040	1.2882	1.2802

Notes: L_μ: number of lags in VAR part μ_x; L_r: number of lags in the ARCH part R_x; L_p: number of lags in the polynomial part $P(z, x)$; $P(z, x)$ is of degree K_z in z with cross product terms exceeding $K_z - I_z$ set to zero; same for K_x and I_x; S_N: negative of the log likelihood divided by sample size (9636); P_Θ: total number of parameters; BIC, HQ, AIC: respectively: Schwartz, Hannan-Quinn and Akaike criterions.

have to formulate a *bivariate* SNP involving both returns and volume to fit a SV model with $\Delta g(t)$ including both series. This is, of course, different from the QMLE setup described in the previous section. A bivariate SNP involving both returns and volume has been estimated by Gallant, Rossi and Tauchen (1992). We relied on an optimal score defined by the following values of the tuning parameters: $\{L_\mu = 2, L_r = 18, L_p = 2, K_z = 4, I_z = 1, K_x = 2, I_x = 1\}$.[13] Since the bivariate SNP involves a very large number of parameters (368), we extended the sample period and considered data on prices and trading volume from a 1928–1987 sample.

[13]For further details, see Gallant, Rossi and Tauchen (1992).

4. EMPIRICAL RESULTS

In this section, we turn our attention to an empirical study of SV models subject to time deformation. As noted in the previous section, the data used consist of the daily closing value of the $S\&P$ composite stock index and the daily volume of shares traded on the NYSE. The data are plotted in Figure 4.1 which consists of two parts: namely, 4.1a displays the return series, while 4.1b contains volume. The empirical results reported here are based exclusively on the series appearing in Figure 4.1. A first subsection will be devoted to SV models not involving the laws of motion of trading volume $\log V_t$. The second subsection will be based on the joint volatility-volume specification.

4.1 Empirical Subordinated SV Models

We fitted two continuous time SV models: the first one contains a simplified volatility equation and is based on the assumption that in a long run of operational time, the Ornstein-Uhlenbeck process is pulled towards zero. The second model corresponds exactly to the setup presented in section 2, where the stochastic volatility is allowed to tend towards any finite level. Two estimation methods, namely the QMLE and EMM presented in the previous section, were applied. The Kalman filter parameter estimates of the zero drift model appear in Table 4.1, while the second volatility process specification is covered in Table 4.2. The EMM estimates appear in Table 4.3.

Table 4.1

Stochastic Volatility with Time Deformation Determined by Past Trading Volume and Prices with Leverage Effects

Sample: 1950–1987, QMLE, nonzero drift

	(1)			(2)			(3)		
	Est	SE	P	Est	SE	P	Est	SE	P
c_v	0.5164	0.9152	0.5726	1.2475	0.3793	0.0010	—	—	—
c_p	–0.2043	0.1734	0.2386	—	—	—	–0.1874	0.0906	0.0386
c_ℓ	0.2461	0.3011	0.4138	—	—	—	0.3333	0.1118	0.0028
Σ	0.0108	0.0034	0.0014	0.0143	0.0037	0.0000	0.0129	0.0045	0.0038
a	–0.0098	0.0040	0.0146	–0.0146	0.0034	0.0000	–0.0105	0.0036	0.0036
b	–0.2163	0.1133	0.0562	–0.2742	0.0853	0.0014	0.1007	0.0752	0.1808
λ	0.0281	0.0109	0.0100	0.0281	0.0109	0.0100	0.0281	0.0109	0.0100
a_1	0.1768	0.0100	0.0000	0.1768	0.0100	0.0000	0.1768	0.0100	0.0000

	(4)			(5)			(6)		
	Est	SE	P	Est	SE	P	Est	SE	P
c_v	1.3210	0.6898	0.0554	—	—	—	—	—	—
c_p	–0.3724	0.0816	0.0000	–0.2252	0.0830	0.0066	—	—	—
c_ℓ	—	—	—	—	—	—	—	—	—
Σ	0.0150	0.0054	0.0052	0.0114	0.0039	0.0032	0.0126	0.0037	0.0008
a	–0.0133	0.0048	0.0052	–0.0117	0.0044	0.0074	–0.0133	0.0037	0.0004
b	–0.1591	0.1168	0.1730	–0.2471	0.0980	0.0118	–0.2379	0.0772	0.0020
λ	0.0281	0.0109	0.0100	0.0281	0.0109	0.0100	0.0281	0.0109	0.0100
a_1	0.1768	0.0100	0.0000	0.1768	0.0100	0.0000	0.1768	0.0100	0.0000

Note: The standard errors reported are based on corrected QMLE asymptotic covariance matrix.

Table 4.2

Stochastic Volatility with Time Deformation Determined by Past Trading Volume and Prices with Leverage Effects

Sample: 1950–1987, QMLE, zero drift

	(1)			(2)			(3)		
	Est	SE	P	Est	SE	P	Est	SE	P
c_v	1.0098	0.6236	0.1054	0.9790	0.7012	0.1626	—	—	—
c_p	−0.1674	0.0959	0.0810	—	—	—	−0.1565	0.0979	0.1098
c_ℓ	0.2562	0.1356	0.0588	—	—	—	0.3542	0.1152	0.0022
Σ	0.0149	0.0051	0.0038	0.0123	0.0035	0.0002	0.0125	0.0034	0.0002
a	−0.0121	0.0040	0.0024	−0.0115	0.0035	0.0008	−0.0107	0.0031	0.0006
λ	0.0281	0.0109	0.0100	0.0281	0.0109	0.0100	0.0281	0.0109	0.0100
a_1	0.1768	0.0100	0.0000	0.1768	0.0100	0.0000	0.1768	0.0100	0.0000

	(4)			(5)			(6)		
	Est	SE	P	Est	SE	P	Est	SE	P
c_v	1.2559	0.5759	0.0292	—	—	—	—	—	—
c_p	−0.3742	0.0767	0.0000	−0.5101	0.0472	0.0000	—	—	—
c_ℓ	—	—	—	—	—	—	—	—	—
Σ	0.0138	0.0037	0.0002	0.0155	0.0058	0.0039	0.0117	0.0033	0.0004
a	−0.0117	0.0032	0.0002	−0.0132	0.0047	0.0026	−0.0113	0.0030	0.0002
λ	0.0281	0.0109	0.0100	0.0281	0.0109	0.0100	0.0281	0.0109	0.0100
a_1	0.1768	0.0100	0.0000	0.1768	0.0100	0.0000	0.1768	0.0100	0.0000

Note: The standard errors reported are based on corrected QMLE asymptotic covariance matrix.

A total of six variants of each model were evaluated, with the sixth being a SV model without time deformation, i.e., imposing $c_\ell = c_v = c_p = 0$. The other five specifications involve time deformation, yet with different functional forms. The most general specification is the unconstrained model with $\Delta g(t)$ as a function of past volume and returns with a leverage effect. The second model involves only volume; the third, only prices with leverage effect; the fourth, prices and volume without leverage; and, finally, the fifth model has $\Delta g(t)$ determined by past price changes. A total of seven coefficients in the zero drift $O-U$ and eight in the other case were estimated. The parameter λ was obtained as a sample average of $\Delta \log y_t$ following the suggestion of Gouriéroux, Monfort and Renault (1993). Moreover, as there appears to be some minor autocorrelation left in $\Delta \log y_t$, we first fitted first-order autoregressive models to $\Delta \log y_t$ and replaced $\Delta \log y_t$ by the residuals to estimate the SV models. The autoregressive coefficients appear as a_1 in both tables. The standard errors reported in Tables 4.1 and 4.2 are based on a QMLE covariance matrix estimator. The EMM method based on simulations allows us to estimate only three out of six variants, namely, the model without the deformation of time and with time deformation either determined by past returns only or returns with leverage effects. The eight parameters (we considered only the nonzero drift volatility specification) were estimated simultaneously, as opposed to the two-step procedure adopted in the QML approach involving the estimation of λ separately. The parameter estimates are presented in Table 4.3, where we compare them to the results of the Kalman filter.

The parameter values all appear to agree with the stochastic process behavior described in section 2. To evaluate the significance of individual coefficients, we rely on the QML t ratios. In particular, the basic continuous time parameters A and Σ are significantly different from zero throughout all specifications and A takes only negative values. The mean coefficient β yields mixed results since it is significant in four out of six specifications. This would mean that we should have a preference for the nonzero drift model if we were to choose between the two volatility specifications. Moreover β always takes negative values, except for the model (3) estimated by EMM, where the long run mean of the volatility process in operational time is much higher than elsewhere. The parameters appearing in the return equation, λ and a_1 are significant, though their values vary depending on the estimation method. In general, the QML

Table 4.3

Stochastic Volatility with Time Deformation Determined by Past Prices with Leverage Effects

Sample: 1950–1987, EMM, nonzero drift

	(3)		(5)		(6)
	EFFGMM	NOISE ADDED	EFFGMM	NOISE ADDED	EFFGMM
c_p	–0.8875	0.5562	–2.2543	0.4292	—
c_ℓ	3.2260	1.5634	—	—	—
c_n	—	0.0101	—	–0.1524	—
Σ	0.0002	0.0001	0.0022	0.0005	0.0159
a	–0.0003	–0.0002	–0.0025	–0.0003	–0.0221
b	0.2173	0.0491	–0.0447	–0.5198	–0.5543
λ	0.0445	0.0447	0.0447	0.0448	0.0529
a_1	0.2125	0.2117	0.2125	0.2127	0.2143
χ^2	113.846	127.082	123.726	129.985	126.563

estimates of λ and a_1 are larger than the EMM estimates in a one-step procedure, while the continuous time parameters A and Σ resulting from the simulation-based method are much lower.

Let us now discuss briefly the estimates of the time deformation parameters c_v, c_p and c_ℓ beginning with the QMLE results. Past volume has a positive impact on $\Delta g(t)$ since c_v takes always positive values. This implies, as noted in section 2, that the marginal effect of increases in trading volume is a volatility process being less persistent in calendar time and more erratic. The leverage coefficient is also positive, while past price change always enter with a negative coefficient in the $\Delta g(t)$ specification. However, since $|c_\ell| > |c_p|$ with $c_\ell > 0$ and $c_p < 0$ it follows that whenever $\Delta \log p_{t-1}$ is negative, we find a greater positive effect of past returns on $\Delta g(t)$ than when $\Delta \log p_{t-1}$ is positive. Hence, bull markets tend to make volatility larger, less persistent and more erratic, while bear markets are associated with a lower volatility with smaller variance. Note that the EMM procedure yields larger values of the $\Delta g(t)$ parameter estimates then QMLE. This is partly due to a different treatment of the time deformation function in the EMM framework, where we did not require $\Delta g(t)$ to average to one in long term as we did in the Kalman filter, but we imposed instead an upper bond of 1.394×10^{65}. Joint tests, based on the QML results, have also been examined. Hence, we complement the Wald tests presented in Tables 4.1 and 4.2 with LR-type tests that appear in Tables 4.4 and 4.5. Tests regarding the time deformation hypothesis appear in the first table. The results indicate that when $\Delta g(t)$ is determined by either one of the individual series, volume or prices the Wald and LR tests are not in agreement and there is also a difference depending on the process specification. However, prices combined with either a leverage effect or trading volume yield robust and strong results supporting significant time deformation. Finally, the three series combined again yield mixed results with the joint LR test favoring time deformation, though none of the coefficients are individually significant for the zero-drift model. In Table 4.5, we turn our attention to a number of LR tests

Table 4.4
Time Deformation Hypothesis Tests (LR)

Series in $\Delta g(t)$	AR(1)	Ornstein-Uhlenbeck
Volume only[a]	2.794	5.492
Prices only[a]	10.599	1.445
Prices with leverage[b]	6.455	8.961
Prices and volume[b]	9.635	11.947
Prices with leverage and volume[c]	15.994	5.299

Note: The likelihood ratio statistic is asymptotically distributed as χ^2 with respectively $a = 1$, $b = 2$ and $c = 3$ degrees of freedom.

Table 4.5
Hypothesis Tests of the Time Deformation Function (LR) —
The Continuous Time AR(1) Model

Hypotheses	AR(1)	Ornstein-Uhlenbeck
$H_0 : \; c_v \neq 0 \;\; c_p = 0 \;\; c_\ell = 0$ $H_A : \; c_v \neq 0 \;\; c_p \neq 0 \;\; c_\ell \neq 0$	13.200	12.911
$H_0 : \; c_v = 0 \;\; c_p \neq 0 \;\; c_\ell \neq 0$ $H_A : \; c_v \neq 0 \;\; c_p \neq 0 \;\; c_\ell \neq 0$	9.538	10.502
$H_0 : \; c_v = 0 \;\; c_p \neq 0 \;\; c_\ell = 0$ $H_A : \; c_v \neq 0 \;\; c_p \neq 0 \;\; c_\ell \neq 0$	5.396	5.877
$H_0 : \; c_v \neq 0 \;\; c_p \neq 0 \;\; c_\ell = 0$ $H_A : \; c_v \neq 0 \;\; c_p \neq 0 \;\; c_\ell \neq 0$	6.359	0.001

regarding the functional specifications of time deformation. We test whether $\Delta g(t)$ is determined by: (1) volume only against the alternative of volume and prices with leverage; (2) prices with leverage only against the same alternative; (3) prices only without leverage; and volume once again against all three series. In each case, the restricted model is rejected. We also test whether leverage should be introduced once prices and volume determine time deformation and find mixed results. In the zero-drift model, we observe a significant leverage effect, while the $O-U$ process appears to have a very flat likelihood surface, making the marginal contribution of leverage to $\Delta g(t)$ negligible.

We turn our attention now to the sample path of the time deformation process $\Delta g(t)$ for a number of specifications. As we could not plot all possible combinations, since it would be quite repetitive, we selected a few representative cases. We first examine the path of time deformation for the AR(1) model with two alternative specifications of $\Delta g(t)$: one involving prices and volume, the other adding leverage effects. Four plots appear in Figure 4.2. Each $\Delta g(t)$ specification yields a pair of plots, one for $\Delta g(t)$, the other for the innovation variance which also depends on $\Delta g(t)$. Figures 4.2a and 4.2b display the patterns of time deformation, both involving prices and volume with leverage effects included in the latter. They appear to be quite similar, though $\Delta g(t)$ with leverage seems to be slightly less erratic. One key feature emerging from both figures, as well as the adjacent plots containing the innovation variance to the volatility process, is the infrequent appearance of sharp peaks in operational time acceleration. Since A is found to negative, this means that the conditional variance function becomes locally extremely erratic, unattached to the previous period and subject to a large variance innovation shock. As noted in the introduction, this finding complements a competing specification of laws of motion via diffusion processes involving jumps. Such processes, proposed by Merton (1967a, b) were built on the premise that one would occasionally observe abnormal information leading to the incidence of a jump in asset prices. Through the time deformation specification, one can view such information arrival as extremely rapid acceleration of market time through the increased trading and price movement per unit of calendar time. The advantage of SV models with time deformation over jump-diffusion processes is that the former might be relatively easier to estimate, at least if one is satisfied with the asymptotically inefficient QMLE algorithm. Indeed, the ML estimation of jump-diffusion processes can be quite involved [see, for instance, Lo (1988) for details].[14]

We turn our attention now to the volatility process itself, i.e., the h_t process as extracted via the Kalman filter procedure. A first caveat to note is that the filtering algorithm we use, like the estimation procedure, is only an approximation of the true latent volatility process. Indeed, the Kalman filtering algorithm ignores all non-Gaussian features of the DGP, as noted in section 3. Jacquier, Polson and Rossi (1992) proposed a procedure that yields an exact extraction algorithm for the volatility process as a by-product of their Bayesian inference procedure for SV models. Their algorithm is numerically quite more involved in comparison to that described in section 3 and is probably not so easy to modify so that a time deformation SV model can be handled [see Ghysels and Jasiak (1993) for further discussion]. Figure 4.3a displays the approximate filter extraction of the volatility process h_t. The figure consists of

[14]There is, of course, a substantial difference between the stochastic process behavior of a jump process and a process with SV having occasionally very large volatility.

two parts, namely, 4.3a displays stochastic volatility as extracted under the assumption of no time deformation. Hence, Figure 4.3a corresponds to a volatility process that one would obtain from the approach proposed by Nelson (1988) and Harvey, Ruiz and Shephard (1994). Figure 4.3b plots h_t extracted from a model with time deformation. In sharp contrast to the standard SV specification, we uncover a very smooth volatility process. This may not be as surprising, given the plots in Figure 4.2 where $\Delta g(t)$ and the innovation variance appeared. Indeed, most of the erratic behavior of h_t obtained through a specification without time deformation is absorbed through the doubly stochastic random coefficient stochastic volatility specification. Once time deformation is taken into account, it appears that the underlying volatility process evolves smoothly in operational time. This yields an alternative interpretation. Indeed, the smooth evolution of h_t in operational time implies that the process is easier to predict over long horizons. This smooth and predictable component appears to be separated from the mote erratic behavior of market time through $\Delta g(t)$. This separation into two components is interesting as it decomposes a volatility process that is itself latent.

Let us now examine further the empirical results obtained from the EMM framework. A goodness of fit test can be performed by computing, a chi-square statistic $N\, m_N'(\tilde{\alpha}_N,\, \tilde{\ominus}_N)\,(\tilde{I}_N)^{-1} m_N\,(\tilde{\alpha}_N,\, \tilde{\ominus})$, which, under a correct specification of the SV model, is asymptotically distributed as χ^2 with degrees of freedom equal to the length of the SNP parameter vector $\ominus$ minus the number of parameters of the SV model, collected in the vector α. All the three models under study failed the chi-square test (see Table 4.3). The value of the test statistic obtained for the nondeformation SV model falls far beyond the critical value. However, the objective function can be significantly improved upon, once we incorporate the time deformation determined by past price changes with leverage effects. A similar but weaker reduction in the value of the statistics can be observed when $\Delta g(t)$ is defined as function of past returns only.

Figure 4.4 shows the improvement in terms of the t-statistics on the scores of the SNP model obtained by fitting a time deformation SV model. The parameters "psi" correspond to the AR coefficients of the SNP model, the "a's" indicate the quadratic and quartic terms, and finally the "tau's" are the coefficients on the ARCH components of the score generator. The time deformation model with leverage effects improves 28 among 34 t ratios, while the model without the leverage performs equally well, reducing 29 out of 34. It is interesting to note here that most of the improved moment fits appear in the "tau" group representing ARCH components. This is of course a key issue, as time deformation should affect the volatility component first and foremost.

In Tables 4.4 and 4.5, we reported joint tests to examine the fit of the SV model with time deformation. In the context of the EMM estimation procedure, we can perform some model diagnostics more directly aimed at the restrictions imposed by the time deformation specification of the volatility equation. Returning to equation (2.10b), we notice that $\Delta g(t)$ controls the intercept AR coefficient as well as the innovation variance. In a simulation context, we can ask ourselves whether we can improve the fit by breaking this link between the correlations and variance of the volatility. We do so by adding an extra term to the equation, namely:

$$(4.1.1) \qquad h_t = [(1 - \exp(A\Delta g(t))]\beta + [\exp(A\Delta g(t)) + c_n\tilde{\nu}_t]h_{t-1} + \nu_t$$

where $\tilde{\nu}_t$ is an i.i.d. $N(0, 1)$ sequence and c_n is an additional parameter. The results in Table 4.3 indicate that the fit deteriorates with the "noise added" specification. We also find larger t statistics on the score vector, is illustrated in Figure 4.5. Out of 34 t ratios, 17 went up in the case of model (3) and 13 out of 34 for model (5). These results suggest that breaking the restrictions obtained in equation (2.10b) do not improve the fit. Also, the "tau" group of moment conditions is the one where the deterioration proclaims itself as one would expect. We simply perturbed the specification by adding a noise term to the AR coefficient. Perturbations in other "directions" may perhaps yield other results. So, far however, we find the time deformation specification the best fit focused so far.

4.2 Empirical Volatility-Volume Models

In this section, we rely exclusively on EMM estimation guided by the bivariate return and volume SNP density described at the end of section 3. The parameter estimates of the bivariate model with time deformation are reported in Table 4.6. A total of 14 parameters are estimated with the empirical score of the bivariate SNP as a guidance of matching moments. In our specification of the time deformation, we did not include the absolute value of returns. We experimented with such a specification as well, but found the model reported in 4.6 a better fit. This does not mean the model does not produce leverage effects. Indeed, the bivariate structure is far more complex than the univariate one and asymmetric responses can arise without it appearing in $\Delta g(t)$. We shall in fact return to this issue shortly.

Judging the adequacy of the bivariate model is no longer as straightforward as judging that of the univariate of course. We mentioned in section 2 that we would therefore analyze the model via impulse response functions. We will follow a strategy proposed by Gallant, Rossi and Tauchen (1993) which consists of computing response

Table 4.6
EMM Parameter Estimates of Bivariate Stochastic Volatility-Volume Model with Time Deformation

μ_P	–0.1762	a_{11}	–0.4550
μ_V	0.9511	a_{12}	–0.4267
a_1	0.4570	a_{21}	0.8026
β_P	–1.1440	a_{22}	–4.6646
β_V	0.9301	Σ_{11}	1.9558
c_P	–1.1204	$\Sigma_{21} = \Sigma_{12}$	–0.1298
c_V	0.3412	Σ_{22}	1.0171

Note: Moment matching using score function of bivariate SNP described by the following turning parameters: $L_\mu = 2$, $L_M = 18$, $L_P = 2$, $K_Z = 4$, $I_Z = 1$, $K_X = 2$, $I_X = 1$.

profiles for the conditional mean and conditional volatility, where the profiles are defined by:

(4.2.1) $$\hat{y}_j(x) = E(y_{t+j} \mid x_t = x) \qquad j = 0, 1, \ldots$$

for the conditional mean profile and:

(4.2.2) $$\hat{V}_j(x) = E[\text{Var}(y_{t+j} \,|\, x_{t+j-1})x_t = x] \qquad j = 0,\ 1,\ \ldots\ .$$

In both cases, y_{t+j} represents a component of the bivariate return-volume process. The conditioning vector x is the one which is perturbed to produce different response profiles. Gallant, Rossi and Tauchen consider three scenarios to compute the response profiles, namely:

(4.2.3) $$x^+ = (y_0',\ y_{-1}',\ y_{-2}',\ \ldots)' + (\delta_y^+,\ 0,\ 0,\ \ldots)$$

(4.2.4) $$x^\circ = (y_0',\ y_{-1}',\ y_{-2}',\ \ldots)'$$

(4.2.5) $$x^- = (y_0',\ y_{-1}',\ \ldots)' + (\delta_y^-,\ 0,\ 0,\ \ldots)$$

where δ_y^+ and δ_y^- are shocks to the nonlinear dynamic system. The vector x° is called the baseline shock while x^+ and x^- are respectively positive and negative shocks. Hence, the conditional profiles are predictions of y_{t+j} for the three different initial conditions listed in (4.2.3) through (4.2.5). The combination of equations (4.2.1) through (4.2.5) yield the conditional mean and volatility profiles $(\hat{y}_j(x),\ \hat{V}_j(x))$ for $x = x^+$, x° and x^-. These response profiles were computed for both stock returns and trading volume. Of course, so far we did not discuss how the conditional means and variances appearing in (4.2.1) and (4.2.2) are obtained. In Gallant, Rossi and Tauchen (1993), the empirical SNP, which yields an estimate of the conditional density was used to compute (4.2.1) and (4.2.2). This empirical density will serve as a benchmark against which we want to measure the success of the bivariate stochastic volatility model with time deformation described in (2.12). In order to compare the impulse response profiles of the empirical SNP with those of the bivariate SV model with time deformation, we produced a simulated sample of data containing 20,000 observations (approximately the size of the empirical data using the estimated model). The simulated data generated by the model were than used to fit an SNP density from which the impulse profiles were computed. To assess the usefulness of the time deformation specification, we also estimated the bivariate stochastic volatility model defined in (2.12) through (2.14) restricting $\Delta g(t) = 1 \forall t$. We followed exactly the same procedure as above and obtained impulse response profiles under a bivariate model specification without time deformation. We have therefore three impulse response profiles to compare, a first from the empirical fit of the SNP from the data, a second from a SNP density fitted to data simulated from a bivariate model with time deformation using the EMM parameter estimates and finally a third which, like the second, is obtained from simulated data, yet without a time deformation specification.

Gallant, Rossi and Tauchen (1993) point out that impulse response profiles for a bivariate volatility-volume process need to take into account the widely documented contemporaneous relationship between price and volume movements [see, e.g., Karpoff (1987) or Tauchen and Pitts (1983)]. From a scatterplot of historical return-volume data, they define three types of impulses which described different scenarios of contemporaneous return-volume shocks. These shocks, called of type A, B and C were constructed to be consistent with the historical range of the data. We do not want to repeat all the details of the computations and findings based on the three types of shocks as they are reported in Gallant, Rossi and Tauchen. Instead, we will single out one particular case which the authors identified as one of the most interesting and novel findings emerging from the impulse response analysis. Namely, Gallant, Rossi and Tauchen found that the leverage effect is essentially a transient effect when analyzed in a bivariate system in sharp contrast to the univariate price shock models which shows a much more persistent wedge between effects of positive δ_y^+ and negative δ_y^- price shocks. We reexamine this issue using both type A and B shocks, the former being a combined price-volume shock while the latter being a pure price shock.[15]

The impulse response profiles are summarized in Figures 4.6a and 4.6b. The former covers shocks of type A while the latter of type B. Each figure has six panels, the top three panels display conditional mean profiles of returns while the lower three panels exhibit conditional variance profiles. Each time, one has (1) impulse responses from the empirical SNP, (2) from the SNP generated by a time deformation bivariate model and (3) one without time deformation. We consider the empirical SNP as being the benchmark case where the data features are summarized. A first observation to make is that the response profiles for the model without time deformation appear to

[15] Shocks of type C relate to volume only, see Gallant, Rossi and Tauchen (1993) for more details, in particular regarding the interpretation given to the three types of shocks.

overstate the volatility responses considerably, both for shocks of type A and B. Besides being off track, we also notice that the baseline shock without time deformation shows a slight kink in both cases which does not appear in the deformation model nor the empirical SNP. While both models appear to confirm the transient nature of the leverage effect, we note the model without time deformation tends to slightly overstate, in relative terms, the initial response of a negative shock. The differences are minor, but the time deformation specification does show an edge, at least on the basis of these impulse response profiles. Other criteria may be found to better discriminate between models, but this we would rather leave for future research.

5. CONCLUSIONS

In this paper, we proposed an empirical class of time deformation stochastic volatility models that were fitted to daily return and volume data for the NYSE. Two estimation procedures were discussed, one involving a Kalman filter QMLE algorithm, the other involving a moment matching principle. A univariate as well as bivariate return-volume model specification were considered.

The framework can easily be extended to deal with high frequency data. For instance, Ghysels and Jasiak (1995) suggest a specification of a time deformed SV models involving arrivals of quotas and bid-ask spreads at 5 and 20 minutes intervals for foreign exchange markets. Last but not least, Ghysels, Gouriéroux and Jasiak (1995) provide a detailed discussion of the stochastic process theory for subordinated processes. They, as well as Conley, Hansen, Luttmer and Scheinkman (1994) discuss various estimation procedures not covered here.

Appendix

Fig 4.1a: S&P 500 (Adjusted)

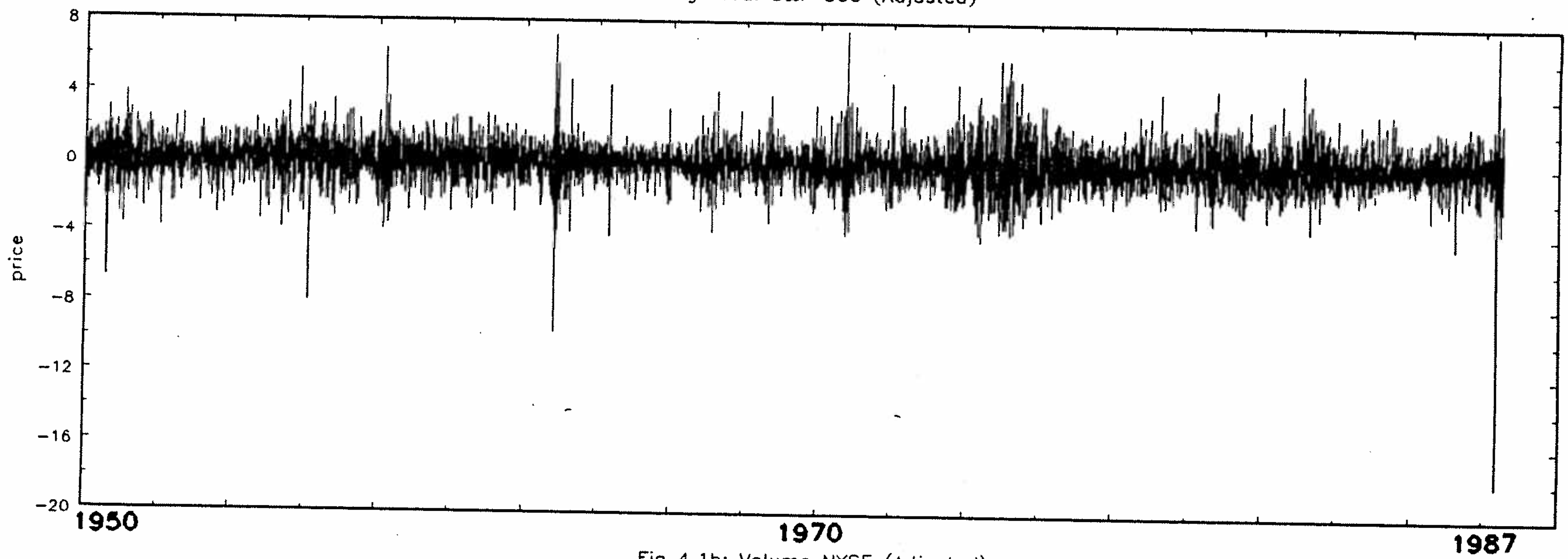


Fig 4.1b: Volume NYSE (Adjusted)

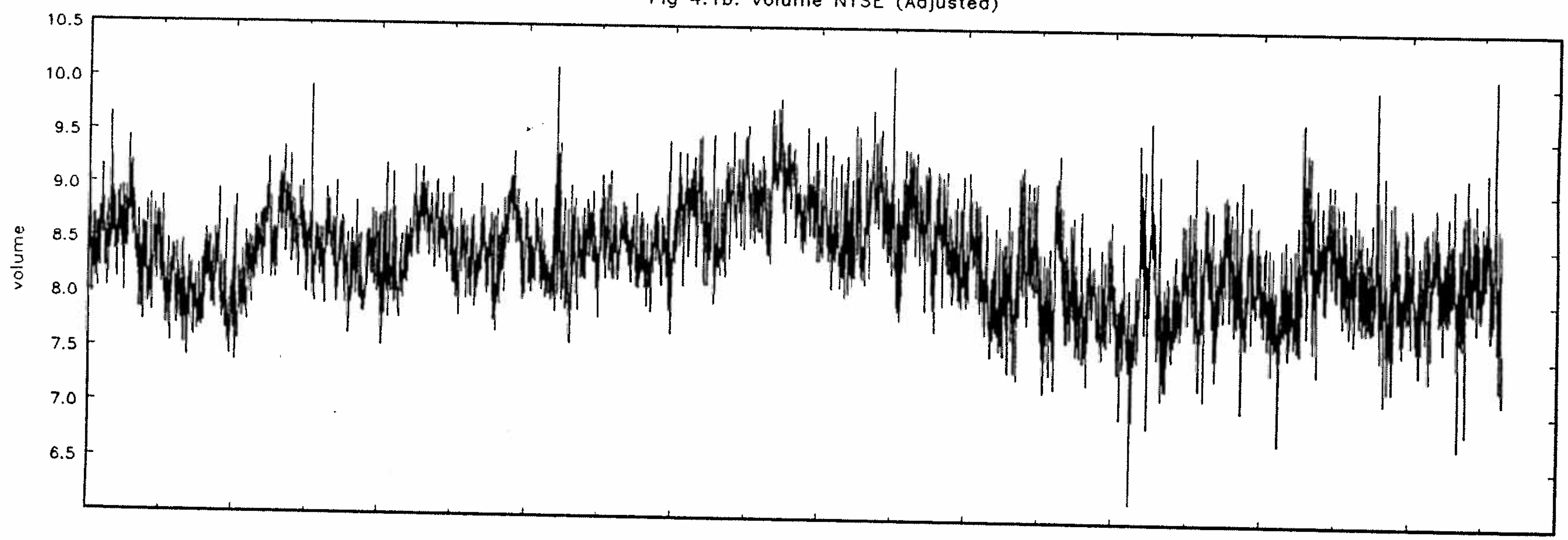


Figure 4.2 : AR(1) SV Model with Time Deformation

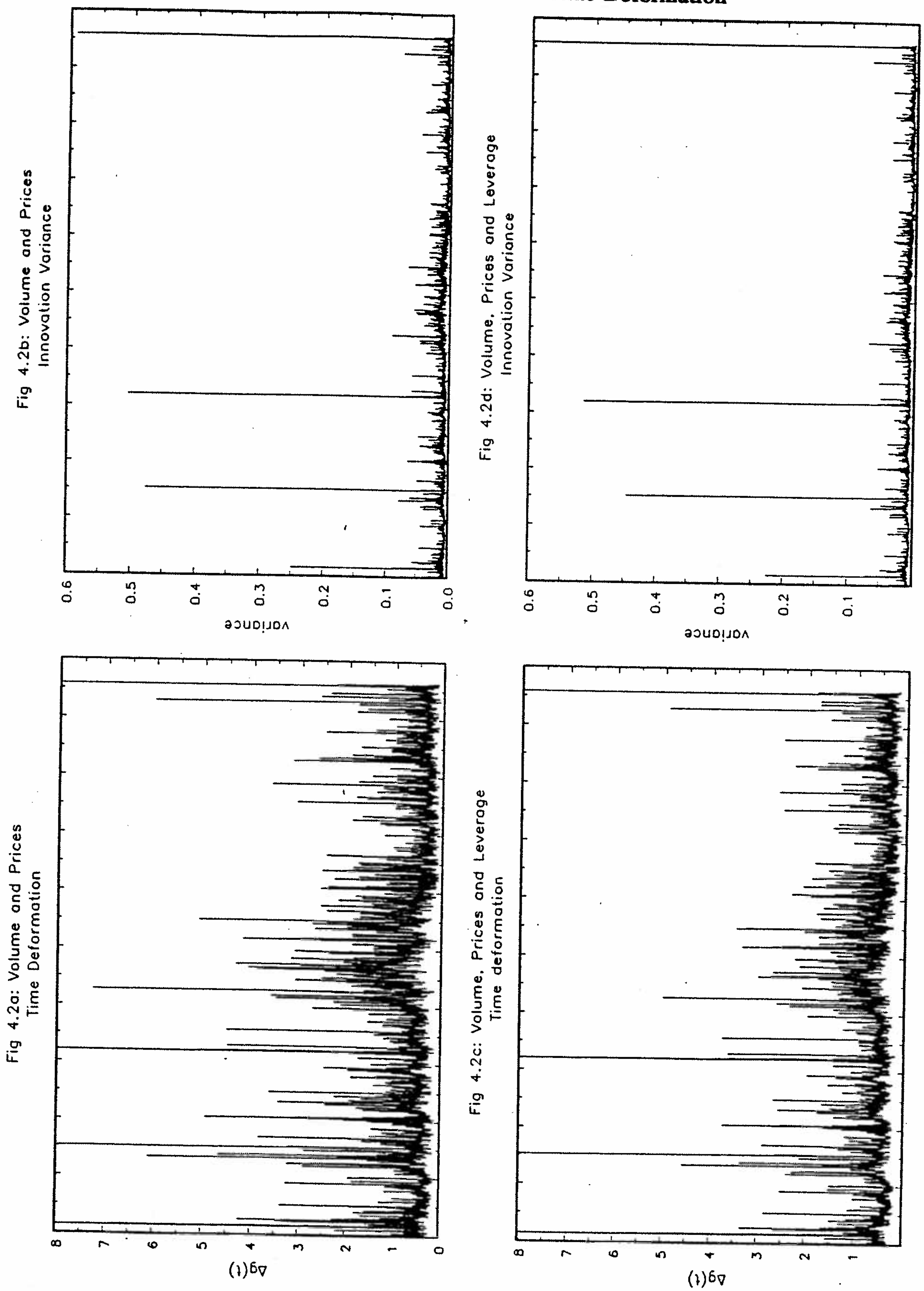


Fig 4.3a: Volatility without Time Deformation

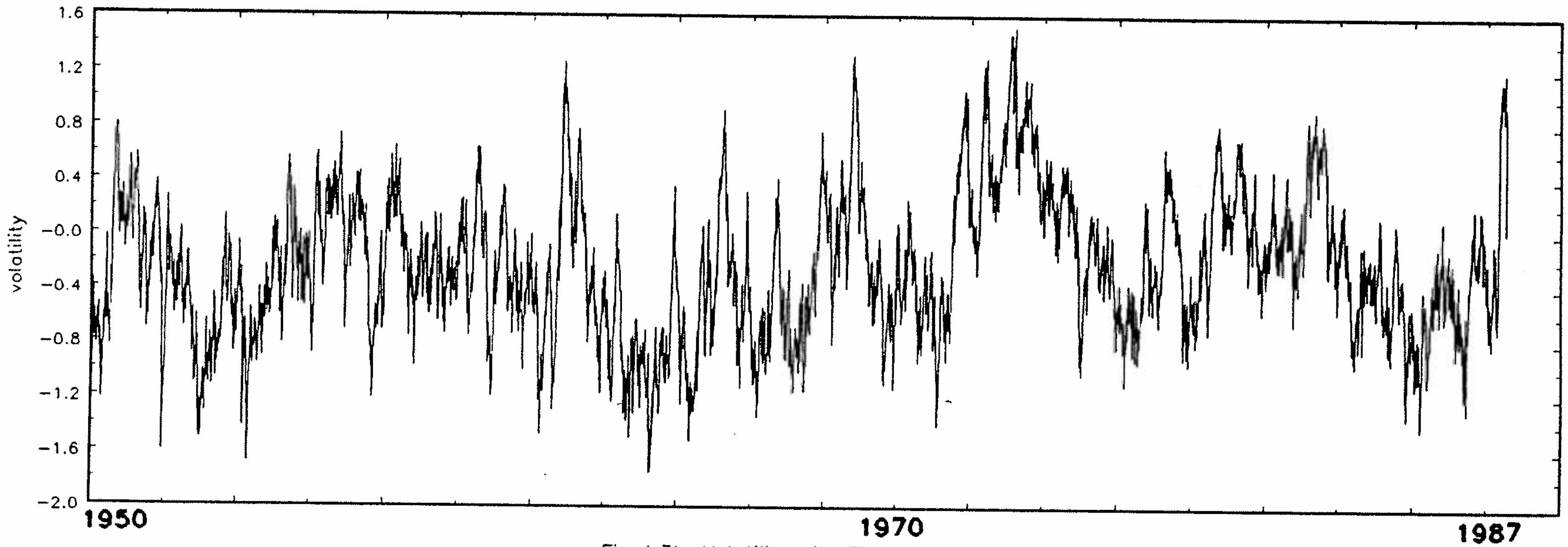
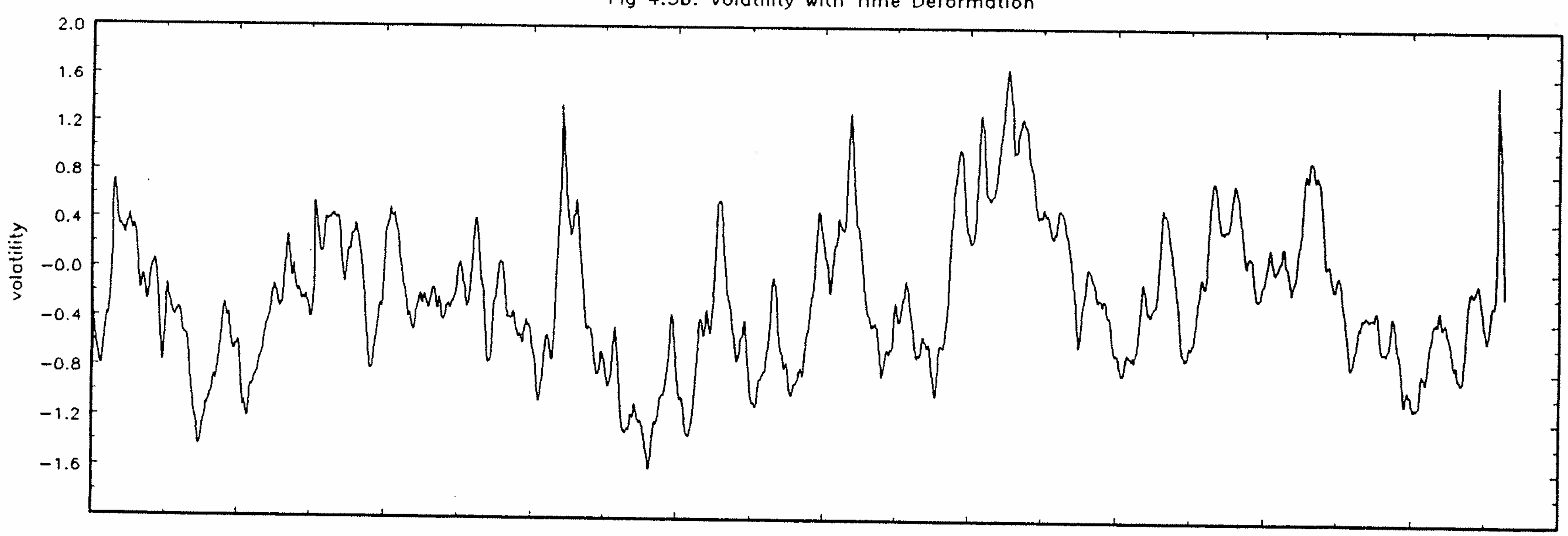


Fig 4.3b: Volatility with Time Deformation



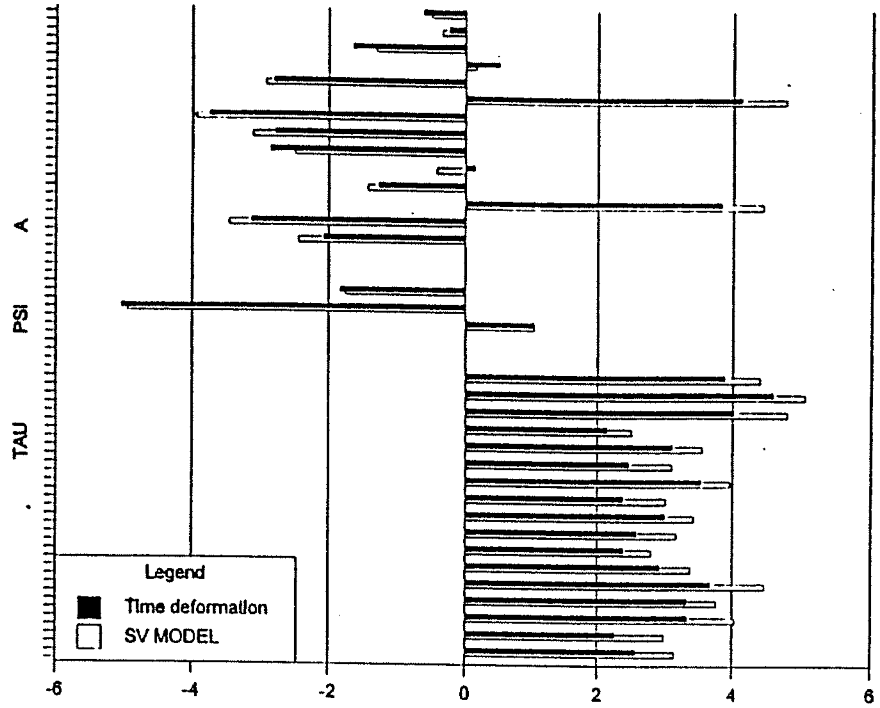


Figure 4.4 : SV MODEL vs TIME DEFORMATION, DETERMINED BY PAST RETURNS AND THEIR ABSOLUTE VALUE

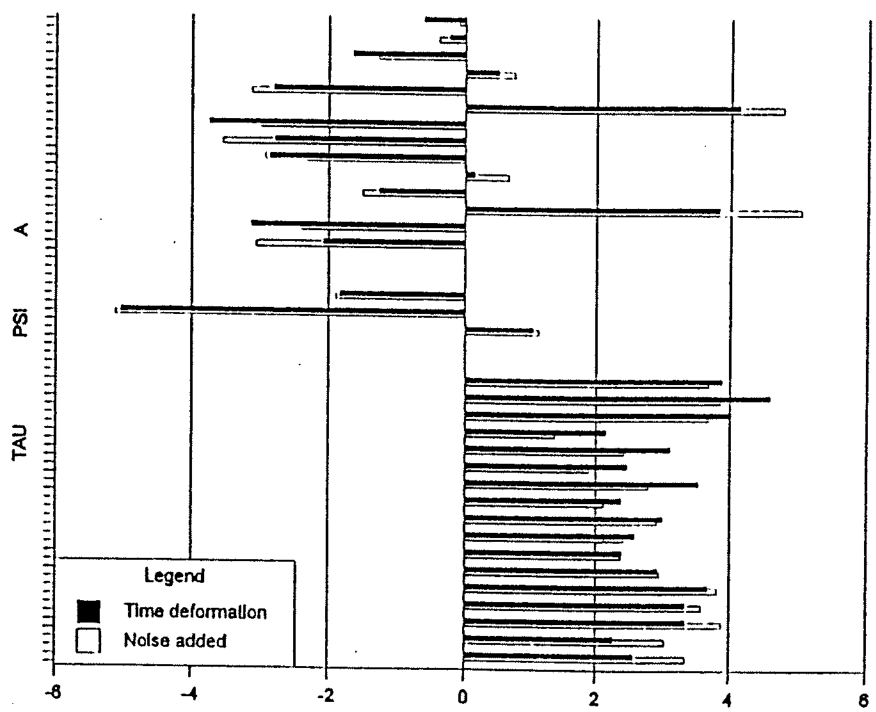


Figure 4.5 : TIME DEFORMATION vs MODEL WITH NOISE ADDED, DETERMINED BY PAST RETURNS AND THEIR ABSOLUTE VALUE

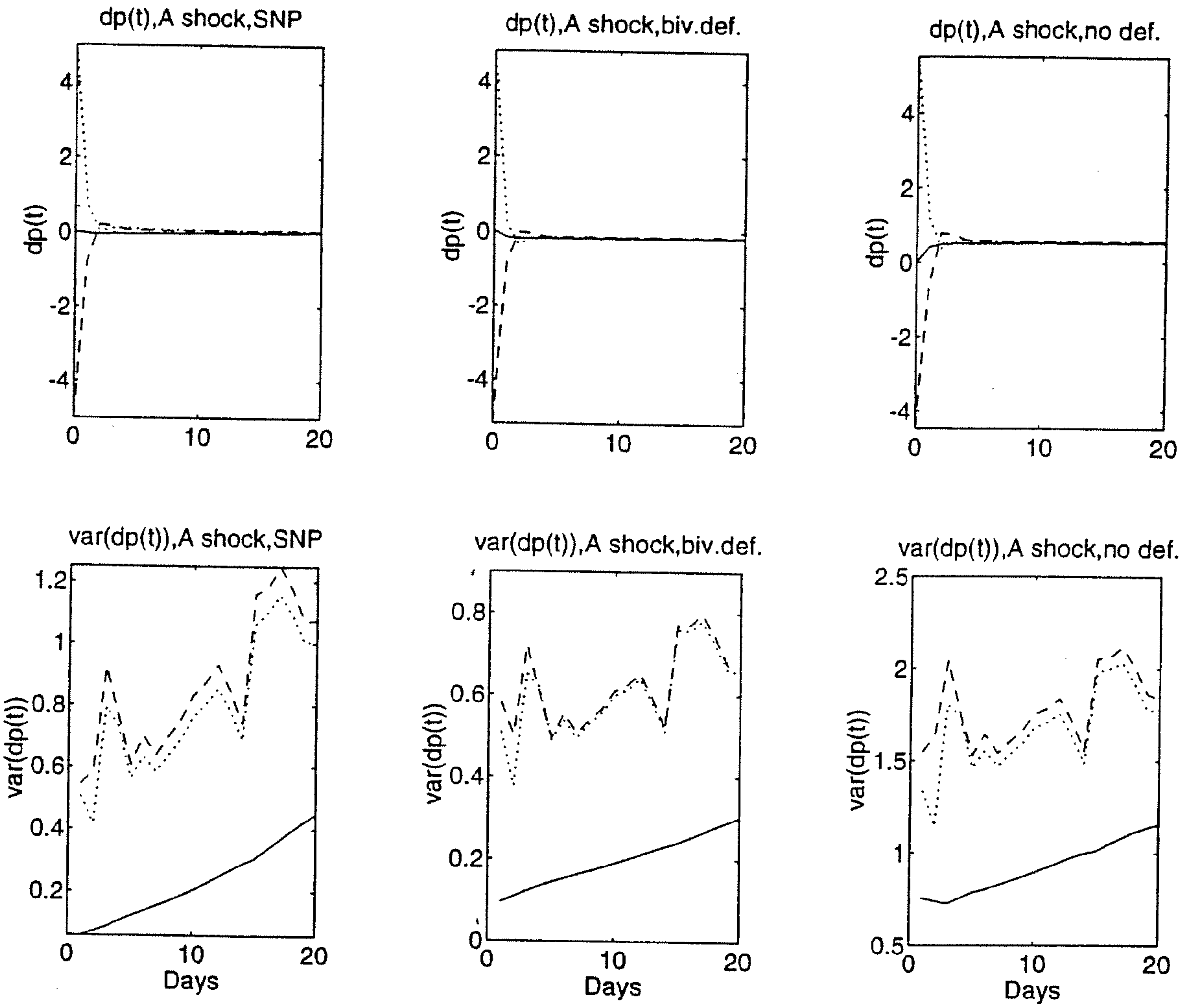


Figure 4.6a

Impulse responses of Δp and volatility to shock type A defined in Gallant, Rossi and Tauchen (1993, p. 892) from bivariate fit of SNP densities to empirical data and simulated data from bivariate SV models with and without time deformation. In each panel, the heavy solid line is the baseline where $\hat{x}_\tau = x_\tau$, the solid line corresponds to a negative A shock where $\hat{x}_\tau = x_\tau + (0, 0, \ldots, \delta y^-_A)'$, and the dashed line corresponds to a positive A shock where $\hat{x}_\tau = x_\tau + (0, 0, \ldots, \delta y^+_A)'$.

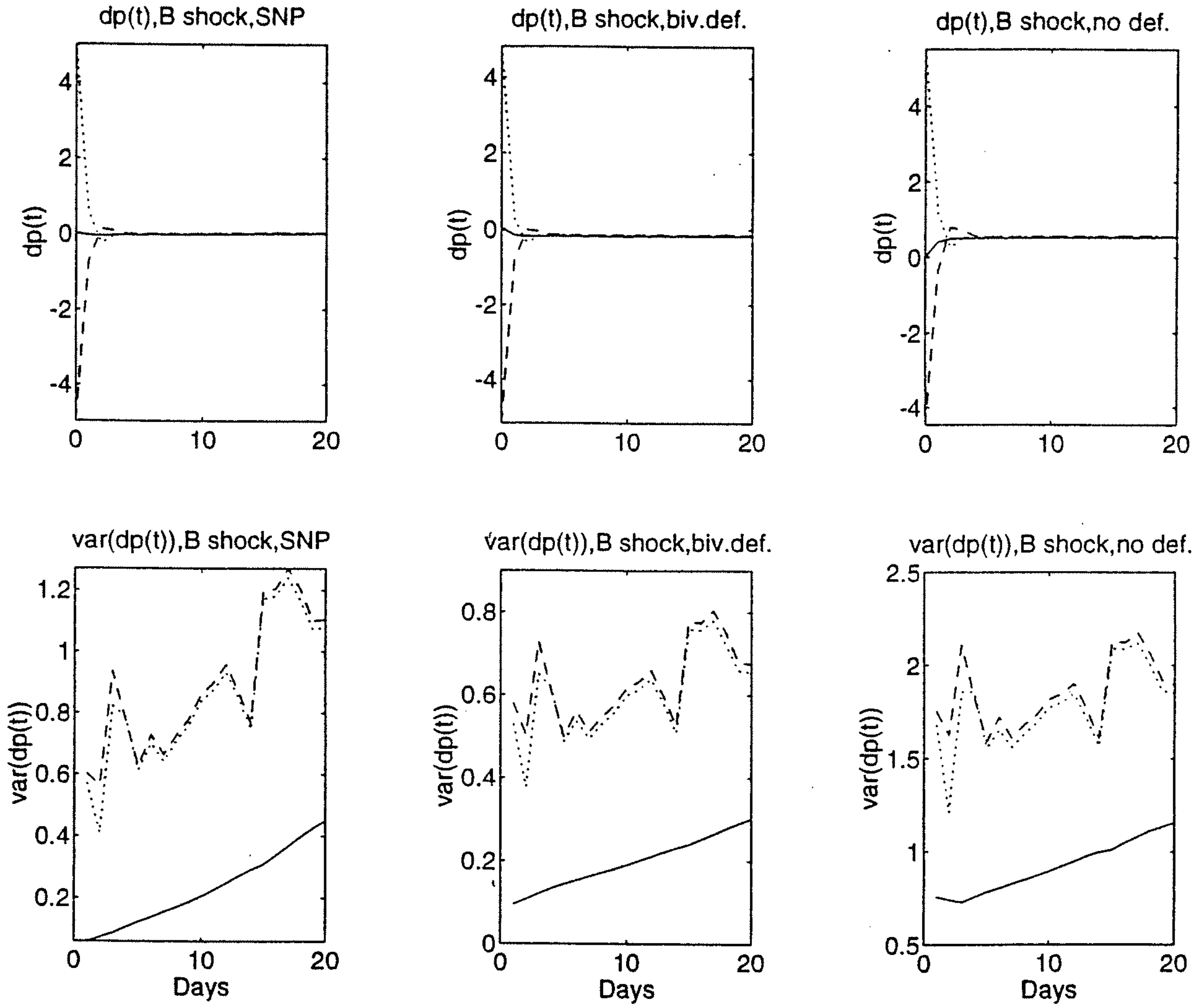


Figure 4.6b

Impulse responses of Δp and volatility to shock type B defined in Gallant, Rossi and Tauchen (1993, p. 892) from bivariate fit of SNP densities to empirical data and simulated data from bivariate SV models with and without time deformation. In each panel, the heavy solid line is the baseline where $\hat{x}_\tau = x_\tau$, the solid line corresponds to a negative B shock where $\hat{x}_\tau = x_\tau + (0, 0, \dots, \delta y_B^-)'$, and the dashed line corresponds to a positive B shock where $\hat{x}_\tau = x_\tau + (0, 0, \dots, \delta y_B^+)'$.

REFERENCES

Andersen, T. and B.E. Sorensen (1994), "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," Discussion Paper, Northwestern University.

Bessembinder, H. and P.J. Seguin (1993), "Price Volatility, Trading Volume and Market Depth: Evidence from Futures Markets," *Journal of Financial and Quantitative Analysis* 28(1), 21–39.

Black, F. (1976), "Studies of Market Volatility Changes," *Proceedings of the American Statistical Association, Business and Economics Section*, 177–181.

Bollerslev, T. and E. Ghysels (1994), "On Periodic Autoregression Conditional Heteroscedasticity," Discussion Paper, CIRANO and *C.R.D.E.*

Bollerslev, T., R.Y. Chou and K.E. Kroner (1992), "Arch Modelling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics* 52(1/2), 5–61.

Bollerslev, T. and R.J. Hodrick (1992), "Financial Market Efficiency Tests," in M.H. Pesaran and M.R. Wickens (eds.), *Handbook of Applied Econometrics, Vol. I — Macroeconomics* (forthcoming).

Chesney, M. and L. Scott (1989), "Pricing European Currency Options: A Comparison of the Modified Black-Scholes Model and a Random Variance Model," *Journal of Financial and Quantitative Analysis* 24, 267–284.

Christie, A. (1982), "The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects," *Journal of Financial Economics* 10, 407–432.

Clark, P.K. (1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica* 41(1), 135–156.

Conley, T., L. Hansen, E. Luttmer and J. Scheinkman (1994), "Estimating Subordinated Diffusions from Discrete Time Data," Presented at the CIRANO-C.R.D.E. conference on Stochastic Volatility, Montréal.

Dacorogna, M.M., U.A. Müller, R.J. Nagler, R.B. Olsen and O.V. Pictet (1993), "A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market," *Journal of International Money and Finance* 12, 413–438.

Detemple, J. and S. Murthy (1993), "Intertemporal Asset Pricing with Heterogeneous Beliefs," Discussion Paper, School of Management, McGill University.

Dimson, E. (ed.) (1988), *Stock Market Anomalies*, Cambridge University Press, Cambridge.

Easley, D. and M. O'Hara (1992), "Time and the Process of Security Price Adjustment," *Journal of Finance* XLVII(2), 577–605.

Foster, D. and S. Viswanathan (1993a), "The Effect of Public Information and Competition on Trading Volume and Price Volatility," *Review of Financial Studies* 6(1), 23–56.

Foster, D. and S. Viswanathan (1993b), "Can Speculative Trading Explain the Volume Volatility Relation," Working Paper, Fuqua School of Business, Duke University.

French, K. (1980), "Stock Returns and the Weekend Effect," *Journal of Financial Economics* 8, 55–69.

French, K. and R. Roll (1986), "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics* 17, 5–26.

French, K.R., G.W. Schwert and R.F. Stambaugh (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics* 19, 3–29.

Gallant, A.R., D.A. Hsieh and G. Tauchen (1991), "On Fitting a Recalcitrant Series. The *Pound/Dollar* Exchange Rate, 1974–83," in William A. Barnett, J. Powell, and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Proceedings of the Fifth Symposium in Economic Theory and Econometrics, Cambridge University Press, Cambridge.

Gallant, A.R., P.E. Rossi, and G. Tauchen (1992), "Stock Prices and Volume," *Review of Financial Studies* 5, 199–242.

Gallant, A.R., P.E. Rossi and G. Tauchen (1993), "Nonlinear Dynamic Structures," *Econometrica* 61, 871–908.

Gallant, A.R. and G. Tauchen (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica* 57, 1091–1120.

Gallant, A.R. and G. Tauchen (1994), "Which Moments to Match?," *Econometric Theory* (forthcoming).

Geman, H. and M. Yor (1993), "Bessel Processes, Asian Options and Perpetuities," *Mathematical Finance* 3, 349–375.

Ghysels, E., C. Gouriéroux and J. Jasiak (1994), "Market Time and Asset Price Movements: Theory and Estimation," Discussion Paper CIRANO and CREST.

Ghysels, E. and J. Jasiak (1993), "Comments on 'Bayesian Analysis of Stochastic Volatility Models'," *Journal of Business and Economic Statistics* (forthcoming).

Ghysels, E. and J. Jasiak (1995), "Trading Patterns, Time Deformation and Stochastic Volatility in Foreign Exchange Markets," Paper presented at the Olsen and Associates HFDF conference, Zürich.

Gladysev, E.G. (1961), "Periodically Correlated Random Sequences," *Soviet Mathematics* 2, 385–388.

Gouriéroux, C. and A. Monfort (1989), "Statistique et modèles économétriques," *Economica*, Paris.

Gouriéroux, C., A. Monfort and E. Renault (1993), "Indirect Inference," *Journal of Applied Econometrics* 8, Supplement, S85–S118.

Hansen, L.P. and T.J. Sargent (1990), "Recursive Linear Models of Dynamic Economies," Manuscript, Hoover Institution, Stanford University.

Harris, L. (1987), "Transaction Data Tests of the Mixture of Distributions Hypothesis," *Journal of Financial and Quantitative Analysis* 22, 127–141.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, C.R. and R.D. Huang (1991), "Volatility in the Foreign Currency Futures Market," *Review of Financial Studies* 4, 543–560.

Harvey, C.R. and R.D. Huang (1992), "Information Trading and Fixed Income Volatility," Unpublished Manuscript, Department of Finance, Duke University.

Harvey, A.C., E. Ruiz and N. Shephard (1994), "Multivariate Stochastic Variance Models," *Review of Economic Studies* 61, 247–264.

Harvey, A.C. and N. Shephard (1993), "The Econometrics of Stochastic Volatility," Discussion Paper, *L.S.E.*

Harvey, A.C. and J.M. Stock (1985), "The Estimation of Higher Order Continuous Time Autoregressive Models," *Econometric Theory* 1, 97–112.

Hausman, J.A. and A.W. Lo (1991), "An Ordered Probit Analysis of Transaction Stock Prices," Working Paper, Wharton School, University of Pennsylvania.

Heston, S.L. (1993), "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies* 6, 327–343.

Huffman, G.W. (1987), "A Dynamic Equilibrium Model of Asset Prices and Transactions Volume," *Journal of Political Economy* 95, 138–159.

Hull, J. and A. White (1987), "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance* 42, 281–300.

Jacod, J. and A.N. Shiryayev (1987), *Limit Theorems for Stochastic Processes*, Springer Verlag, New York.

Jacquier, E., N.G. Polson and P.E. Rossi (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics* 12, 371–388.

Johnson, H. and D. Shamro (1987), "Option Pricing When the Variance Is Changing: *Journal of Financial and Quantitative Analysis* 22, 143–152.

Karpoff, J. (1987), "The Relation Between Price Changes and Trading Volume: A Survey," *Journal of Financial and Quantitative Analysis* 22, 109–126.

Lakonishok, J. and S. Smidt (1988), "Are Seasonal Anomalies Real? A Ninety Year Perspective," *Review of Financial Studies* 1, 403–425.

Lamoureux, C. and W. Lastrapes (1990), "Heteroscedasticity in Stock Return Data: Volume versus GARCH Effect," *Journal of Finance* XLV(1), 221–229.

Lamoureux, C. and W. Lastrapes (1992), "Forecasting Stock Return Variance: Toward an Understanding of Stock Implied Volatilities," Working Paper, J.M. Ohlin School of Business, Washington University.

Lamoureux, C. and W. Lastrapes (1993), "Endogenous Trading Volume and Momentum in Stock Return Volatility," Working Paper, J.M. Ohlin School of Business, Washington University.

Lo, A.W. (1988), "Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data," *Econometric Theory* 4, 231–247.

Madan, D.B. and F. Milne (1991), "Option Pricing with V.G. Martingale Components," *Mathematical Finance* 1, 39–55.

Madan, D.B. and E. Seneta (1990), "The V.G. Model for Share Market Returns," *Journal of Business* 63, 511–524.

Melino, A. (1991), "Estimation of Continuous Time Models in Finance," Working Paper, University of Toronto.

Melino, A. and S.M. Turnbull (1990), "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics* 45(1/2), 239–267.

Merton, R. (1973), "An Intertemporal Capital Asset Pricing Model," *Econometrica* 41, 867–888.

Merton, R.C. (1976a), "Option Pricing When Underlying Stock Returns Are Discontinuous," *Journal of Financial Economics* 3, 125–144.

Merton, R.C. (1976b), "The Impact on Option Pricing of Specification Errors in the Underlying Stock Price Returns," *Journal of Finance* 31, 333–350.

Merton, R.C. (1990), *Continuous–Time Finance*, Basil Blackwell Inc., Cambridge, MA.

Nelson, D.B. (1988), "Time Series Behavior of Stock Market Volatility and Returns," Ph.D. dissertation, MIT.

Nelson, D.B. (1989), "Modelling Stock Market Volatility Changes," *Proceedings of the American Statistical Association, Business and Economics Section*, 93–98.

Nelson, D.B. (1990), "ARCH Models as Diffusion Approximations," *Journal of Econometrics* 45(1/2), 7–39.

Nelson, D.B. (1991), "Conditional Heteroscedasticity in Asset Returns: A New Approach," *Econometrica* 59, 347–370.

Pagan, A. and G.W. Schwert (1990), "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, Cambridge University Press, Cambridge.

Potter, S.M. (1991), "Nonlinear Impulse-Response Function," Discussion Paper, U.C.L.A.

Richardson, M. and T. Smith (1992), "A Direct Test of the Mixture of Distributions Hypothesis. Measuring the Information Flow Throughout the Day," Working Paper, Wharton School, University of Pennsylvania.

Sawyer, K.R. (1993), "Continuous Time Financial Models: Statistical Applications of Stochastic Processes," in G.S. Maddala et al. (eds), *Handbook of Statistics — Volume 11: Econometrics*, North-Holland Publications Co., Amsterdam.

Schwert, G.W. (1990), "Indexes of U.S. Stock Prices from 1802 to 1987," *Journal of Business* 63, 399–426.

Scott, L. O. (1987), "Option Pricing When the Variance Changes Randomly: Theory, Estimation and an Application," *Journal of Financial and Quantitative Analysis* 22, 419–438.

Stein, E.M. and J. C. Stein (1991), "Stock Price Distributions with Stochastic Volatility: An Analytic Approach," *Review of Financial Studies* 4, 727–752.

Stock, J.H. (1988), "Estimating Continuous Time Processes Subject to Time Deformation," *Journal of the Statistical Association* 83(401), 77–84.

Tauchen, G.E. and M. Pitts (1983), "The Price Variability Volume Relationship on Speculative Markets," *Econometrica* 51(2), 485–505.

Tjøstheim, D. (1986), "Some Doubly Stochastic Time Series Models," *Journal of Time Series Analysis* 7, 51–73.

Wang, J. (1993), "A Model of Competitive Stock Trading Volume," Working Paper, MIT.

Wiggins, J. (1987), "Options Values Under Stochastic Volatility: Theory and Empirical Evidence," *Journal of Financial Economics* 19, 351–372.

Zadrozny, P. (1988), "Gaussian Likelihood for Continuous Time ARMAX Models When Data Are Stocks and Flow at Different Frequencies," *Econometric Theory* 4, 108–124.

Trading Patterns, Time Deformation and Stochastic Volatility in Foreign Exchange Markets

1. INTRODUCTION

The interbank FX market for foreign exchange transactions is one of the prime examples of the recent trends in globalization of trading in international financial markets. The Bid/Ask prices quoted by various firms and banks are recorded over the 24 hours per day and displayed worldwide by news services, such as Reuters or Telerate. Due to the overlapping periods of activity of market makers located over the 3 continents — America, Europe and Asia, a sequential pattern in intra-day trading is observed. Works by Wasserfallen and Zimmerman (1985), Wasserfallen (1989), Feinstone (1987), Ito and Roley (1987), Müller et al. (1990), Goodhart and Figliouli (1991), Bollerslev and Domowitz (1993) and Dacorogna et al. (1993) are examples of a growing interest in research in this area. Most of them document and examine both daily and weekly seasonalities in the volatility of foreign exchange rates.

The seasonal phenomena in the volatility of foreign exchange markets can be modelled in a variety of ways. One possibility to accommodate seasonality is to modify the traditional ARCH or GARCH type models.[1] Another strategy is to seasonally adjust the data, a practice quite common for economic time series but which is not without its longstanding controversies.[2] Alternatively, the market volatility can be tied to the intensity of trading via a subordinate stochastic process representation, as suggested by Clark (1973). This approach has been adopted in some recent works by researchers from Olsen and Associates [see, for example, Dacorogna et al. (1992, 1993), Müller et al. (1992)]. Instead of modelling asset price behavior in calendar time, price movements can be represented as being driven by an information arrival process which itself evolves randomly yet with certain predictable patterns through time. Formally, daily returns, $x(\Delta t) = \log(p(t)/p(t-1))$, are hence redefined as $\log p(t)/p(t-1) = x(\Delta g(t))$ where $g(t)$ is a positive, increasing stochastic process, sometimes called directing process. This setup can be referred to as time deformation since the relevant time scale is no longer calendar time t but *operational* time $g(t)$. Let us point out some advantages of this approach. As emphasized by Mandelbrot and Taylor (1967), it easily accommodates leptokurtic distributions for asset returns. Clark (1973) has shown that within this framework, comovements between trading

[1]ARCH models with seasonality are discussed in Bollerslev and Ghysels (1994).

[2]See Ghysels (1994) and Miron (1994) for further discussion as well as Andersen and Bollerslev (1994) for applications.

volume and asset returns can easily be modelled. Finally, time deformation yields a random variance equivalent to a *stochastic volatility* model. These ideas have been refined and extended in several ways for foreign exchange markets. Dacorogna et al. (1993) proposed time scales related to a measure of worldwide activity, based on an empirical scaling law of returns relating the mean absolute change of the logarithmic middle price to calendar time. It is intuitively based on the notion that as the world market time "slows down", depending on the number of markets active and on their local intra-day pattern, price volatility decreases and vice versa. Since this time deformation concept is based on *average* market activity at any point in time, it accommodates the repetitive seasonal pattern. Dacorogna et al. (1993) do not fully exploit, however, the framework of subordinated processes suggested by Clark as they forego the information in the *current* market activity. In this paper we adopt the generic framework of Ghysels and Jasiak (1994) and propose a stochastic volatility model with time deformation which blends features of an average *and* a conditional market activity.

The empirical work is based on the data provided by Olsen and Associates. The series consist of DEM/USD, JPY/USD and JPY/DEM exchange rates and contain all quotes that appeared on the interbank Reuters network over the entire year from October 1, 1992 through September 29, 1993. Although this data bank contains a bid and ask price for each quote along with the time to the nearest even second, several researchers (see, for example, Dacorogna et al. (1993), Müller et al. (1993), Moody and Wu (1994)) consider a single price series constructed as a logarithmic average of asks and bids. In section 2 we examine the data and discuss the advantages and shortcomings of this approach. The stochastic volatility model in its generic form is presented in section 3. In section 4, we discuss observable stochastic processes which approximate the market activity and appear in our specification of operational time. In section 5 we report the empirical estimates of the stochastic volatility model with time deformation based on intra-day market activity. Section 6 concludes the paper.

2. MARKET DYNAMICS AND THE DISTRIBUTIONAL PROPERTIES OF ASKS AND BIDS

In this section we provide a statistical analysis of asks and bids and study the behavior of their geometric average. We examine the descriptive statistics, the autocorrelation patterns and also investigate marginal and joint empirical densities. Since the outcomes of time scale adjustments are for us of primary concern, the data are both analyzed on a real time (tick-by-tick) basis, and over a fixed 20 minute sampling interval.

The high frequency data consist of interbank FX price quotes for three exchange rates: the Deutschmark/US Dollar (DEM/USD), the Japanese Yen/US Dollar (JPY/USD) and the Yen/Deutschmark (JPY/DEM) rate. The numbers of observations in the three samples are, respectively, 1,472,266, 570,839 and 159,004. Although the ask and bid sequences are reported simultaneously for every transaction, a vast majority of researchers study a single price series constructed as a logarithmic average of asks and bids. Following the notation adopted by Dacorogna et al. (1993), the returns on the foreign exchange market are thus defined as:

$$\begin{aligned}\Delta x(t) &= x(t) - x(t-1) \\ &= \tfrac{1}{2}[(\ell og\, \text{ask}(t) + \ell og\, \text{bid}(t)) - (\ell og\, \text{ask}(t-1) + \ell og\, \text{bid}(t-1))]\,,\end{aligned}$$

or,

$$\begin{aligned}\Delta x(t) &= \tfrac{1}{2}[(\ell og\, \text{ask}(t) - \ell og\, \text{ask}(t-1)) + (\ell og\, \text{bid}(t) - \ell og\, \text{bid}(t-1))] \\ &= \tfrac{1}{2}[\Delta \ell og\ \text{ask}(t) + \Delta \ell og\ \text{bid}(t)]\,.\end{aligned}$$

Usually it is assumed that the dynamics of the $x(t)$ series reflect the general pattern of market activity. One could argue, however, that the logarithmic middle price averages out outcomes of distinct trading strategies of buyers and sellers. Indeed, the real time data reveal several differences between asks and bids. Table 2.1 presents the summary statistics of $\Delta \ell og$ ask(t) and $\Delta \ell og$ bid(t), the two components of $\Delta x(t)$, as well as of $\Delta x(t)$ compared across markets in real time. Table 2.2 contains the same statistical summary over a fixed 20 minute interval of time scale. We report the mean, variance, standard deviation, skewness coefficient, excess kurtosis (i.e., the empirical kurtosis –3), the minimum and maximum values as well as the range. A 95% confidence interval of the mean and variance estimators are also provided.

In real time, we find in general that $\Delta \ell og$ ask(t) has a higher mean and a larger variance than $\Delta \ell og$ bid(t). However, the first two moments of asks and bids differ only marginally as compared to the discrepancies reported in moments of order 3 or 4. In fact, the most relevant differences arise in terms of asymmetry and tail properties. On the JPY/USD and JPY/DEM markets, the ask series are skewed to the right, while the bids are skewed to the left. The quotes on the DEM/USD exchange rates are both skewed to the right and show little differences in absolute values of the skewness coefficients. On the contrary, on the JPY/DEM market, we report a 434 times higher absolute value of the skewness coefficient of asks compared to bids. More excess kurtosis is found in the ask series as well. The difference is either slight, as it is the case of the most active and hence, most regularly behaved DEM/USD market, moderate in the JPY/USD quotes, where excess kurtosis in asks is almost 3.5 times higher than in bids or extreme on the JPY/DEM market where the ask coefficient is almost 1523 times larger than the excess kurtosis of the bid series.

Two observations can be made regarding the third and fourth moment statistics reported in tables 2.1 and 2.2. First, the differences in skewness and kurtosis for bids and asks in the tick–by–tick data indicate that there are far more extreme changes in the ask quotations than there are in the bids. As noted before, these differences are particularly important for the JPY/DEM and JPY/USD markets. A second observation is with respect to the comparison of the kurtosis statistics obtained from real time and twenty minute sampling. In Ghysels, Gouriéroux and Jasiak (1995) it is shown that for a time deformed process $X(\Delta g(t))$ there is an increase in kurtosis due to time deformation when the mechanism generating $\Delta g(t)$ is independent of X. This would yield larger excess kurtosis for the twenty minute sampled series in comparison with the tick–by–tick series. The results in tables 2.1 and 2.2 show that this is the case for the DEM/USD series and for the bid series of the JPY/USD market. All other series do not have this feature.

The logarithmic middle price seems to follow the asymmetric pattern of the bid quotes, both in terms of the sign and the magnitude of skewness coefficients. The thickness of tails in the $\Delta x(t)$ series appears also to be determined rather by bids than by asks at least on those markets where the largest bid-ask discrepancies in terms of excess kurtosis were reported, i.e., JPY/USD and JPY/DEM.

The descriptive statistics resulting from data sampled over 20 minute intervals, presented in Table 2.2, provide us some insights on the time scale adjustment effects. The results in Tables 2.1 and 2.2 indicate that the sampling scheme has an immediate and very strong impact on the distributional properties of the data. We report largely different values of the first four moments of quotes on the same exchange rates sampled on the adjusted time scale.[3] Besides, data show much less variety across the markets in a sense that the basic statistics defining the distinct character of the 3 data sets become much less dissimilar. It seems that on the aggregated time scale, some of the properties identifying the individual series are getting attenuated. Accordingly, we do not observe either the bid-ask discrepancies, at least to the extent reported in the real time. For this reason, statistics on both quote sequences and their logarithmic average appear more coherent as well.

To visualize the differences between the $\Delta \ell og$ ask(t), $\Delta \ell og$ bid(t) and $\Delta x(t)$ series in terms of their distributional properties, we present plots of the corresponding empirical univariate densities. (See figure 2.1, appendix 1.) For clarity of exposition, we cover only one market, namely JPY/DEM featuring extreme bid-ask discrepancies in real time.

Figures 2.2–2.3 display the bivariate distributions of [$\Delta \ell og$ ask(t), $\Delta \ell og$ bid(t)], the univariate distributions of the two series, as well as the contour plots of quotes recorded both in real time and on the adjusted time scale. A typical shape of the bivariate density can be described as a sudden, very pronounced peak surrounded by some smaller ones within a large domain of infrequently quoted values. In all data sets, the empirical densities are stretched out along one axis of the ellipse, indicating a strong positive correlation between $\Delta \ell og$ ask(t) and $\Delta \ell og$ bid(t). The shapes shown on the contour plots confirm a higher variance of data sampled at 20 minute intervals and suggest more correlation between both quote sequences on the 20 minute grid.

The issue that remains to be investigated is whether the distributional properties revealed by quotes recorded over one year are shared by samples over shorter time horizons, like one month or one day. A closely-related problem is the stability of the empirical densities through calendar time versus presence of some seasonal or irregular patterns.

We selected 6 monthly subsamples consisting of quotes recorded in October and

[3] This phenomenon has been documented for series aggregated from daily to weekly or to monthly sampling frequencies [see Drost and Nijman (1993)].

December 1992 as well as in January, March, May and July 1993. We analyzed both the tick-by-tick data and quotes sampled at 20 minute intervals. For simplicity of exposition, we report the results for one market only, i.e., the JPY/USD. (See figures 2.4–2.5, appendix 1.)

The variety of shapes of the bivariate empirical densities of monthly subsamples throughout the year reveals the complexity of seasonal phenomena in exchange rates. In the monthly tick-by-tick data sets, asks and bids show less discrepancies than in the entire sample. The JPY/USD market is particular in a sense that $\Delta \ell og$ ask(t) still take values over a larger range than $\Delta \ell og$ bid(t), has a higher variance and longer tails. (See figures 2.4, appendix 1.) Especially ask quotes on the JPY/USD rates remain symmetric in October, while bids display a strong skewness. Interestingly, the October asymmetry is common to all bid sequences and exhibited also by asks on the JPY/DEM exchange rates.

The density of the average JPY/USD price, $\Delta x(t)$, seems, in general, to take on values over the bid's range. However, it does not reflect the bid's asymmetry. On the remaining markets where the ask and bid densities are more similar in terms of range and variance, skewness in the $\Delta x(t)$ sequence seems to be determined by the skewness of bids.

Quotes on the JPY/USD rates, sampled at 20 minute intervals, do not reveal the "October skewness". (See figure 2.5, appendix 1.) Instead, an asymmetry in the asks' density is observed in January, while bids display asymmetric behavior either in January, May, July and December. The bids and asks prices recorded on the remaining two markets have similar asymmetric distributions in almost all monthly subsamples. Apparently, on the adjusted time scale, every month has its own particular rhythm and pattern of trading, as reflected by a characteristic tail behavior.

Since asks and bids on DEM/USD and JPY/DEM rates exhibit on the 20 minute grid similar distributional properties, the general tendency of these markets is expected to be well-approximated by $\Delta x(t)$. In case of the JPY/USD quotes, their logarithmic average mimics the tail behavior of bids rather than asks.

The empirical densities of daily samples are shown in figures 2.6–2.7, appendix 1. Our data consists of quotes recorded over 4 days of the week of October 5 through 12, 1992. Daily patterns are examined on the most active DEM/USD market. The empirical densities of daily subsamples differ again in terms of shape. In the tick-by-tick records, Wednesday's and Sunday's distributions are more stretched out and are more symmetric than the Monday's and Friday's ones. On Monday, both asks and bids are characterized by long right tails, while on Friday the asymmetry is exhibited by bids only. The middle price $\Delta x(t)$ reflects again the distributional properties of bids.

The quotes sampled over the 20 minute grid show a variety of daily densities although due to a small number of observations, Sunday's data can be disregarded. As we have observed in the monthly data, the adjustments of the time scale imparts asymmetries and thus seasonality on days (months) where they are not reported in the tick-by-tick data. For example, thick and uneven Wednesday's tail reappear on all markets.

Many of the seasonal phenomena can also be uncovered within the autocorrelation patterns of the series. As this issue will be discussed in section 4, we concentrate on serial dependencies on the real and the adjusted time scales up to lag 100.

The autocorrelations in bids and asks on both time scales are persistent and do not reveal any new facts. Although we do not report the cross-correlation functions, some insights on the price dynamics are worth being presented. As we have inferred from the empirical densities, asks and bids sampled at 20 minute intervals are more correlated than asks and bids in the real time. In fact, covariances of data on the adjusted time scale are almost equal to one on all markets and vary between 0.7–0.8 across markets in the tick-by-tick samples. In terms of the lagged dependence, the first tick is of primary importance for the ask and bid price adjustments. The cross-correlation at lag 1 is negative and varies on the markets between –0.2 and –0.3. The cross-correlations drop dramatically within the next tick indicating still a significant, although a very low, positive dependence (less than 0.03) at lag 2. At higher lags, the dependence between ask and bid series remains extremely low and occasionally takes on significant values. On the 20 minute scale, the real time cross-correlations sum up to one significant lag observed on all markets of a negative value close to –0.1.

To investigate the persistence in $(\Delta x(t))^2$, modelled within the SV framework, we computed the autocorrelation functions of $(\Delta \ell og \text{ ask}(t))^2$, $(\Delta \ell og \text{ bid}(t))^2$ and $(\Delta x(t))^2$. On all markets, squared values of returns, asks and bids in real time show similar, persisting patterns of serial dependence. (See figure 2.8, appendix 1.) The same behavior is revealed by data sampled at 20 minute intervals on the DEM/USD and JPY/USD markets. The squares of returns $(\Delta x(t))^2$ on the JPY/DEM exchange rates are exceptional, as they do not follow the autocorrelation pattern of squared values of ask and bids (see figure 2.9, appendix 1).

The time scale adjustments have shown, so far, either to alleviate some extremes in the distributional structure of the tick-by-tick data or to impart some phenomena related to the seasonality unobserved in real time. Two interpretations seem to be plausible: by sampling at fixed time intervals, we either extract the necessary information out of the noisy tick-by-tick records and reveal the essential properties of the data, or we forego important information and hence obtain an oversimplified image of the true underlying processes. The evidence we have presented, indicates that, apart from some exceptions, the behaviors of asks and bids, the two components of the logarithmic middle price, are much more coherent on the 20 minute time scale. Hence, the middle price incrèments $\Delta x(t)$ approximate better the general tendency of the quotes. By choosing the 20 minute grid to model volatility in the $\Delta x(t)$ series, we need to make the necessary adjustments in the traditional SV model to accommodate several aspects of seasonality. In the next section we explain how this can be achieved by modelling the stochastic volatility within a time deformation framework.

Table 2.1: Summary Statistics (real time (tick-by-tick) data)

DEM/USD

	mean	var	std. dev.	skewness	excess kurtosis	min	max	range
$\Delta\,\ell og\,\mathrm{ask}(t)$	0.992E–07 (–0.345E–06, 0.543E–06)	0.754E–07 (0.752E–07, 0.755E–07)	0.00027	0.04352	5.4334	–0.00603	0.00972	0.01576
$\Delta\,\ell og\,\mathrm{bid}(t)$	0.900E–07 (–0.337E–06, 0.517E–06)	0.697E–07 (0.696E–07, 0.699E–07)	0.00026	0.05181	4.5265	–0.00660	0.00923	0.01584
$\Delta\,x(t)$	0.992E–07 (–0.302E–06, 0.501E–06)	0.617E–07 (0.615E–07, 0.618E–07)	0.00024	0.05438	6.0310	–0.00616	0.00947	0.01565

JPY/USD

	mean	var	std. dev.	skewness	excess kurtosis	min	max	range
$\Delta\,\ell og\,\mathrm{ask}(t)$	–0.218E–06 (–0.119E–05, 0.760E–06)	0.141E–06 (0.140E–06, 0.141E–06)	0.00037	0.02470	21.9921	–0.00944	0.00939	0.01885
$\Delta\,\ell og\,\mathrm{bid}(t)$	–0.233E–06 (–0.117E–05, 0.705E–06)	0.130E–06 (0.129E–06, 0.130E–06)	0.00036	–0.00158	6.5750	–0.00907	0.00916	0.01824
$\Delta\,x(t)$	–0.217E–06 (–0.109E–05, 0.655E–06)	0.112E–06 (0.112E–06, 0.113E–06)	0.00033	–0.00174	10.7756	–0.00921	0.00925	0.01847

JPY/DEM

	mean	var	std. dev.	skewness	excess kurtosis	min	max	range
$\Delta\,\ell og\,\mathrm{ask}(t)$	–0.260E–03 (–0.344E–03, –0.175E–03)	0.293E–03 (0.291E–03, 0.296E–03)	0.01714	167.02110	31359.7910	–0.03059	3.21939	3.24999
$\Delta\,\ell og\,\mathrm{bid}(t)$	–0.166E–05 (–0.338E–05, 0.481E–07)	0.121E–06 (0.120E–06, 0.122E–06)	0.00034	–0.38500	20.5896	–0.00975	0.00995	0.01971
$\Delta\,x(t)$	–0.169E–05 (–0.337E–05, –0.236E–07)	0.115E–06 (0.114E–06, 0.116E–06)	0.00033	–0.39260	18.8470	–0.00968	0.00662	0.01631

Table 2.2: Summary Statistics (20 minute sampling interval)

	mean	var	std. dev.	skewness	excess kurtosis	min	max	range
DEM/USD								
$\Delta \ell og\, \text{ask}(t)$	0.769E–05 (–0.547E–05, 0.208E–04)	0.846E–06 (0.829E–06, 0.864E–06)	0.00092	0.2974	9.0182	–0.00841	0.01344	0.02186
$\Delta \ell og\, \text{bid}(t)$	0.757E–05 (–0.554E–05, 0.207E–04)	0.840E–06 (0.824E–06, 0.858E–06)	0.00091	0.2048	9.1872	–0.00842	0.01345	0.02187
$\Delta x(t)$	0.770E–05 (–0.533E–05, 0.207E–04)	0.830E–06 (0.813E–06, 0.847E–06)	0.00091	0.2496	9.3027	–0.00842	0.01344	0.02186
JPY/USD								
$\Delta \ell og\, \text{ask}(t)$	–0.663E–05 (–0.200E–04, 0.681E–05)	0.873E–06 (0.856E–06, 0.891E–06)	0.00093	0.17402	12.6636	–0.00923	0.01097	0.02021
$\Delta \ell og\, \text{bid}(t)$	–0.663E–05 (–0.198E–04, 0.655E–05)	0.841E–06 (0.824E–06, 0.858E–06)	0.00091	0.01485	10.2305	–0.00924	0.01000	0.01925
$\Delta x(t)$	–0.664E–05 (–0.197E–04, 0.641E–05)	0.824E–06 (0.807E–06, 0.841E–06)	0.00090	0.09053	10.9400	–0.00924	0.01049	0.01973
JPY/DEM								
$\Delta \ell og\, \text{ask}(t)$	–0.161E–04 (–0.313E–04, –0.100E–05)	0.100E–05 (0.979E–06, 0.102E–05)	0.00100	–0.06358	7.3136	–0.00988	0.01143	0.02132
$\Delta \ell og\, \text{bid}(t)$	–0.162E–04 (–0.311E–04, –0.126E–05)	0.972E–06 (0.951E–06, 0.993E–06)	0.00098	–0.22131	7.1505	–0.01003	0.01028	0.02032
$\Delta x(t)$	–0.161E–04 (–0.311E–04, –0.123E–05)	0.963E–06 (0.949E–06, 0.990E–06)	0.00098	–0.13480	7.1586	–0.00996	0.01086	0.02082

3. STOCHASTIC VOLATILITY AND TIME DEFORMATION

In this section we provide a brief summary of the stochastic volatility model with time deformation presented in Ghysels and Jasiak (1994). Following the work by Hull and White (1987), Johnson and Shanno (1987), Scott (1987), Wiggins (1987), Chesney and Scott (1989), Stein and Stein (1991) and Heston (1993), we call a stochastic volatility model the following set of equations:

$$dy(t) = \mu y(t)\,dt + \sigma(t)\,y(t)\,dW_1(t)\ , \tag{3.1a}$$

$$d\log\sigma(t) = a(b - \log\sigma(t))\,dt + c dW_2(t)\ , \tag{3.1b}$$

where $W_1(t)$ and $W_2(t)$ are two independent, standard Wiener processes. Ghysels and Jasiak (1994) suggested to adopt the framework of equations (3.1a) and (3.1b) and define the volatility process as a subordinated stochastic process evolving in a time dimension set by market activity. This approach has been motivated by the works of Mandelbrot and Taylor (1967) as well as Clark (1973). The complex and quite frequently irregular behavior of asset prices becomes simpler and hence easier to model once we assume that the volatility is tied to some observed or unobserved variables, like the information arrival, which determine the dynamics of trades.[4] Hence, we assume that there exist an operational time scale of the volatility process, with $s = g(t)$, a mapping between operational and calendar time t, such that:[5]

$$dy(t) = \mu y(t)\,dt + \sigma(g(t))\,y(t)\,dw_1(t)\ , \tag{3.2a}$$

$$d\log\sigma(s) = a(b - \log\sigma(s))\,ds + c dw_2(s)\ . \tag{3.2b}$$

Following Stock (1988), we use the notation $g(t)$ for the directing process to indicate some *generic* time deformation, which may include trading volume besides many other series that help to determine the pace of the market. Before discussing what might determine $g(t)$, we would like to make some observations regarding equations (3.1a) and (3.1b). Indeed, it should first be noted that the equations collapse to the usual stochastic volatility model if $g(t) = t$. Obviously, there are several possible specifications of $\sigma(g(t))$. Moreover, one could correctly argue that defining volatility

[4]The microfoundations for time deformation and the process of price adjustments can be found most explicitly in Easley and O'Hara (1992).

[5]The mapping $s = g(t)$ must satisfy certain regularity conditions which will be discussed later.

as a subordinated process amounts to suggesting a more complex law of motion in comparison to the Ornstein-Uhlenbeck (henceforth, O-U) specification appearing in (3.1b). This interpretation is valid, yet it should be noted that, through $g(t)$, one can associate many series other than the security price $y(t)$ to explain volatility; hence, one implicitly deals with a multivariate framework. Moreover, as we have pointed out, the time deformation setup enables us to handle rather complex structure through the subordinated representation.

To enhance our understanding of the mechanism of the process, we first consider the system (3.2) in its continuous and discrete time versions. To simplify the presentation, let us set $b = 0$ and discuss a continuous time AR(1). An investor's information can be described by considering the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and the nondecreasing family $F = \{\mathcal{F}_t\}_{t=0}^{+\infty}$ of sub-σ-algebras in calendar time. Furthermore, we let Z_t be a m-dimensional vector process adapted to the filtration F, i.e., Z_t is $\mathcal{F}_t$-measurable. The increments of the time deformation mapping g will be assumed to be $\mathcal{F}_{t-1}$ measurable via the logistic transformation:

$$\frac{dg(\tau; Z_{t-1})}{d\tau} \equiv g(\tau; Z_{t-1}) \equiv \exp(c'Z_{t-1}) / \left\{ \frac{1}{T} \sum_{t=1}^{T} \exp(c'Z_{t-1}) \right\} , \tag{3.3}$$

for $t-1 \leq \tau < t$.[6] Equation (3.3), setting the speed of changes in operational time as a measurable function of calendar time process Z_{t-1}, is completed by the additional identification assumptions:

$$0 < g(\tau; Z_{t-1}) < \infty , \tag{3.4}$$

$$g(0) = 0 , \tag{3.5}$$

$$\frac{1}{T} \sum_{t=1}^{T} \Delta g(t) = 1. \tag{3.6}$$

These three conditions guarantee that the operational time clock progresses in the same direction as calendar time without stops or jumps.[7] Given that g is constant between consecutive calendar time observation via (3.3), its discrete time analogue

[6]The fact that the denominator in (3.3) contains a sample average may suggest that $\sigma(g(t))$ is not measurable with respect to the filtration $\mathcal{F}_t$ in calendar time. However, the denominator in (3.3) is there for reasons of numerical stability of the algorithms. Since it is only a scaling factor, its presence is of no conceptual importance.

[7]See Stock (1988) for a detailed discussion of the identification assumptions.

$\Delta g(t) \equiv g(t) - g(t-1)$ takes the same logistic form appearing in (3.3). At this point, we will not present the components of the Z_{t-1} vector. As we will discuss this issue in the next section, let us just indicate that, in principle, Z_{t-1} consists of any process related to the information arrival. Ghysels and Jasiak (1994) show that the solution in calendar time can be expressed as:

$$\Delta \log y_t - a_1 \Delta \log y_{t-1} - \lambda = e^{h_t} \varepsilon_t , \tag{3.7}$$

$$h_t = [(1 - \exp(A\Delta g(t)))]\beta + \exp(A\Delta g(t)) h_{t-1} + v_t , \tag{3.8}$$

$$v_t \sim N(0, -\Sigma(1 - \exp(2A\Delta g(t)))/ - 2A) . \tag{3.9}$$

Equations (3.7) and (3.8) constitute the basic set of equations for the discrete time representation of the SV model with a subordinated volatility process which evolves at a pace set by $\Delta g(t)$. A linear state-space representation of the system (3.7)–(3.8) can be estimated by maximizing the conditional maximum likelihood function within the Kalman filter framework. Following Harvey, Ruiz and Shephard (1994), we rewrite equation (3.7) as:

$$\log[\Delta \log y_t - a_1 \Delta \log y_{t-1} - \lambda]^2 = h_t + \log \varepsilon_t^2 , \tag{3.10}$$

where: $E \log \varepsilon_t^2 = -1.27$ and $Var \log \varepsilon_t^2 = \pi^2/2$. Defining $\varsigma_t = \log \varepsilon_t^2$, we obtain:

$$\log[\Delta \log y_t - a_1 \Delta \log y_{t-1} - \lambda]^2 = -1.27 + h_t + \varsigma_t . \tag{3.11}$$

Apart from the parameter λ, whose treatment is discussed for instance by Gouriér-oux, Monfort and Renault (1993), the coefficients of this state-space model are time-varying and, hence, similar to the specification proposed by Stock (1988), except for the properties of the ς_t process which is no longer Gaussian. Consequently, the estimation procedure based on the Kalman filter will result here in a quasi-maximum likelihood estimates, as pointed out by Harvey, Ruiz and Shephard (1994). The details of the QMLE algorithm for time deformed SV models are discussed in Ghysels and Jasiak (1994); while Ghysels, Gouriéroux and Jasiak (1994) present a detailed account of subordinated process theory and their estimation.

4. DIRECTING PROCESSES FOR MARKET ACTIVITY

The model structure described in the previous section is a generic one where, apart from some regularity conditions and the logistic form, the specification of $\Delta g(t)$ was left open. Clark (1973), Tauchen and Pitts (1983) and Ghysels and Jasiak (1994) studied stock returns and used a time deformation model with trading volume as proxy for market activity. It is well known that for foreign exchange markets trading volume is difficult to obtain. Hence, we need to consider other series. The *Olsen and Associates* data base provides several possibilities to model market activity. The purpose of this section is to discuss the different approaches one could consider.

Our strategy will consist of distinguishing "regular" or average market activity, and deviations from the expected level of activity. For example, when European financial markets open and start active trading in say the DEM/USD currency exchange, each market participant has a certain expectation of the number of quotes arriving during the first five minutes, the next five minutes, and so on. Some mornings, trading is more brisk or even sometimes frenzy-like. On other mornings, the market activity is down relative to its usual rhythm. Every part of the trading day has a certain reference norm of activity against which one portrays the latest quote arrivals. What is true for quote arrivals also holds for other market indicators like bid-ask spreads, returns, absolute value of returns, etc. The model specification strategy which we will adopt is to incorporate into $\Delta g(t)$ measures of "regular" or average market activity and series representing deviations from average trading patterns. To continue with the quote arrival example, we can formulate $\Delta g(t)$ as:

$$\Delta g(t) = \exp(c' Z_{t-1}) \equiv \exp(\Theta_{qa} nqa_{t-1} + \Theta_{qd}(nqa_{t-1} - nq_{t-1})) \ , \qquad (4.1)$$

where the scaling constant appearing in (3.3) has been omitted from (4.1). Hence, from (4.1) we have that: $Z_{t-1} = (nqa_{t-1}, (nq_{t-1} - nq_{t-1}))$ where nqa_{t-1} is the mean number of quotes arriving over the interval $t-1$, while nq_{t-1} is the actual number of quotes which arrived in $t-1$.

To clarify this, let us consider the plots appearing in figure 4.1 in appendix 2. The figures consist of six plots, the left side displaying graphs with results from data sampled at 5 minute intervals and the right panel containing the 20 minute sampling frequency equivalent. We study the three markets of the Olsen data set, namely

DEM/USD, JPY/DEM and JPY/USD. Each plot covers a span of a week, omitting the weekends, and displays the average number of quotes, computed over the entire sample, for each 5 (left) or 20 (right) minute time intervals of the week. The plots display the repetitive intra-day cycle which is so typical of high frequency exchange rate date. The 5 minute plots are, of course, more jagged than the 20 minute ones, but each shows clearly the patterns of quote arrivals repeating each 24 hour cycle. The graphs displayed in figure 3.1 represent the nqa_{t-1} series used to model $\Delta g(t)$. The number of quote arrivals is one candidate series to measure market activity, besides other series which we shall discuss shortly.

Before turning to these other series, it is worth drawing attention to a special case of time deformation. Suppose for the moment that $\Theta_{qa} = 0$ in (4.1). Then, $\Delta g(t)$ is purely a function of the repetitive daily pattern of $\{nqa_t\}$ which amounts to volatility being a periodic autoregressive process:

$$h_t = \gamma_t + \alpha_t h_{t-1} + W_t \ , \tag{4.2}$$

where γ_t and α_t are changing every 5 or 20 minutes, depending on the sampling frequency, with a 24 hour repetitive cycle, i.e., $\gamma_t = \gamma_s$, $\alpha_t = \alpha_s$ with $s = t + 24$ hours.[8] A periodic model like (4.2) resembles the class of periodic ARCH processes proposed by Bollerslev and Ghysels (1994) in analogy with periodic ARMA models for the mean which have been extensively studied. Of course, the parameter variation in (4.2) is determined by $\gamma_t = (1 - \exp(A\Delta g(t)))$ and $\alpha_t = (A\Delta g(t))$.

We noted that quote arrivals are not the only measure of market activity, and indeed several other series in the Olsen data file could be considered. Figure 4.2 of appendix 2 displays the intra-daily pattern of bid-asks spreads. The figure has the averages computed on a weekly basis of the average bid-ask spreads during 5 or 20 minute intervals. We notice in figure 4.2 a reasonably regular 24 hour pattern for bid-ask spreads but by far not as pronounced and regular as the quote arrivals displayed in figure 4.1 of appendix 2. Following the example in (4.1), we can formulate a directing process as follows, using the same principle:

$$\Delta g(t) \equiv \exp(\Theta_{sa} spa_{t-1} + \Theta_{sd}(spa_{t-1} - sp_{t-1})) \ , \tag{4.3}$$

[8]Since the averages nqa_t were computed on a weekly basis, there might be some slight differences from one day to the next one over an entire week. Yet, judging on the basis of figure 4.1, those differences appear minor.

where spa_{t-1} is the sample average computed on a weekly basis of the mean spread over the interval $t-1$ while sp_{t-1} is the mean spread actually realized.

Last, but certainly not least, we can use absolute return. The weekly averages are displayed in figure 4.3 of appendix 2 this time. The absolute return series has been used by Müller et al. (1990) to model an activity scale. It was observed by these authors that absolute returns exhibited clear structures reflecting market activity through the repetitive cycle of business hours. Indeed, we recover such a pattern, as in figures 4.1 and 4.2, in the case of absolute returns, although the patterns are again not as regular as in the case of quote arrivals. If we were to use only absolute returns, we could construct a directing process:

$$\Delta g(t) \equiv \exp(\Theta_{ra} ara_{t-1} + \Theta_{rd}(ara_{t-1} - ar_{t-1})) \,, \tag{4.4}$$

where ara_{t-1} is the sample average of absolute returns while ar_{-1} are the actual realization for time interval $t-1$.

Whatever measure suits best to formulate the directing process is ultimately an empirical question of model specification and diagnostics. In (3.1) through (3.4) we took each of the series separately in their expected value format and deviations from the mean. However, one could easily combine the series and create a generic directing process:

$$\begin{aligned}\Delta g(t) \equiv \; & \exp(c' Z_{t-1}) \equiv \exp(\Theta_{qa} nqa_{t-1} + \Theta_{sa} spa_{t-1} + \Theta_{ra} ara_{t-1} \\ & + \Theta_{qa}(nqa_{t-1} - nq_{t-1}) + \Theta_{sd}(spa_{t-1} - sp_{t-1}) + \Theta_{rd}(ara_{t-1} - ar_{t-1}))\,.\end{aligned} \tag{4.5}$$

A priori one should expect that the formulation in (4.5) has a lot of redundancy, particularly with respect to the nqa, spa and ara time series. Presumably, the best representation is to pick one of the averages as representative and to measure market activity as a combination of the selected average process plus the series measuring deviations from regular market activity. The latter could be measured either by one, two or all three series entering (4.5).

This is precisely the modelling strategy which we will adopt in the next section. Before discussing the estimation results, we conclude, however, with a discussion of the time series properties of the $(nqa_t - nq_t)$, $(spa_t - sp_t)$ and $(ara_t - ar_t)$ series, i.e., the series measuring deviations from regular market activity. We will do this through the autocorrelation function for each series. These autocorrelations are plotted in figures 4.4 through 4.6 of appendix 2. They are complementary to the plots of the weekly averages. The first of three figures covers the ACF of the $\{nqa_t - nq_t\}$ series. We notice a very strong and repetitive pattern in all three markets. This means that average quote arrivals, as displayed in figure 4.1 of appendix 2 are not the only source of periodic patterns appearing in equation (3.1), but also the deviations from market average is strongly autocorrelated with seasonal patterns. When we turn our attention to figure 4.5 of appendix 2, which covers the bid-ask spread series $\{spa_t - sp_t\}$, we observe less seasonal autocorrelation, at least on a daily basis, but still observe a weekly lag. Since weekends were deleted to compute these autocorrelation functions, one recovers positive autocorrelation at around 360 lags. This weekly pattern is present for both the DEM/USD and JPY/USD market. For the JPY/DEM market, we recover a daily seasonal pattern, however, quite similar to that of figure 4.4. Finally, we turn our attention to the absolute return market deviation series $\{ara_t - ar_t\}$ in figure 4.6. Unlike the two previous series, we find no particular regular patterns in the ACF. Instead, we find a slowly decaying pattern starting from a first order autocorrelation which is much higher than the previous ones, namely .25 instead of around .05 for those appearing in figures 4.4 and 4.5. Even after 800 lags, we still have an autocorrelation above .05.

5. EMPIRICAL RESULTS

There are three currency exchange markets available in the Olsen and Associates data set, we will devote a subsection to each market. We begin with the most active DEM/USD market which is covered in section 5.1 followed by JPY/DEM and finally, JPY/USD markets which are covered in section 5.2. Before discussing the actual results, a few observations are in order regarding estimation. It was noted in the previous section that the details of the QMLE algorithm are omitted here as they appear in Ghysels and Jasiak (1994). The numerical optimization of the quasi-likelihood function was accomplished via simulated annealing. The algorithm, which is described in Goffe et al. (1994), appeared to be the best equipped to deal with the multiplicity of local maxima which tricked most other conventional algorithms we tried. Also, for reasons of numerical stability, we rescaled the quote and spread series by 1.e–03 while the absolute return series was rescaled by 1.e–01.

5.1 The DEM/USD Foreign Exchange Market

In section 4 we noted that our model strategy for formulating the mapping between calendar time and operational time would consist of picking one of the three series measuring anticipated market activity and combine it with the set of series reflecting deviations from averages. Table 5.1 reports the estimation results obtained from the 20 minute sampling interval for three model specifications, each involving different measures of average market activity, as appearing in equations (4.1) through (4.5) complemented with several combinations of the deviations from average market activity. To avoid reporting too many empirical results, we present models with *nqa* and *ara* variables of average market activity and omit those with *spa* which yield quite similar results. The three surprise terms appear either simultaneously or separately in models summarized in table 5.1. Besides the point estimates, we also report standard errors which were computed using a heteroskedasticity consistent QMLE covariance matrix estimator. One should recall that the QMLE procedure is asymptotically inefficient, yet the standard errors in Table 5.1 reveal that all series entering $\Delta g(t)$, no matter what specification is used, appear significant. Hence, the standard errors do not give us much guidance on what model specification to pick. Before elaborating further on model choice, let us discuss first the interpretation of the estimates. One should first of all note that all the coefficients Θ_{ij} have negative signs. Obviously, each coefficient measures a partial effect. However, from microstructure models we know, for instance, that as the time interval between quotes decreases, one expects spreads to increase (see, for instance, Easley and O'Hara (1992) for further discussion). Hence, each of the series reflect movements that are obviously not unrelated. The mapping between calendar time and operational time we investigate is, of course, one based on statistical fit. Let us distinguish first the coefficients related to average market activity from those related to deviations from the normal pace. The first example covers average quotes. Negative coefficients Θ_{qa} and a imply that during times that average quotes are high, market volatility becomes more persistent and less erratic. Obviously, high quote arrivals do not necessarily reflect a high information content, but often it means many markets are active simultaneously. Comparing figures 4.1 and 4.2, we note that high average quote arrivals appear to be associated with higher bid-ask spreads, at least for the DEM/USD market discussed here. Likewise, comparing figures 4.1 and 4.2, we note the same thing for absolute returns, at least again for the DEM/USD market.

The coefficients related to deviations from normal market activity are also negative. Since deviations are measured as average minus actual realizations, it is clear that, with negative a coefficient, above normal market activity increases volatility and vice versa. Also, operational time increases (decreases) when market activity is above (below) average. It must also be noted that each specification of $\Delta g(t)$ involved lagged values of the deviations from market activity. This was, of course, done in order to guarantee that $\Delta g(t)$ is based on variables that are measurable with regard to $t-1$ information. From the autocorrelation functions in figures 4.4 through 4.6, we also know, however, that the first order autocorrelations for each of the market activity deviation processes are positive.

With all entries being significant for the 20 minute DEM/USD specifications, we must rely on other criteria to discriminate among models. In the remainder of this section, we will focus on the models appearing in the first column of tables 5.1 through 5.3. These models contain all three measures of deviations from average market activity combined with each of the three measures of average market activity.

We first turn our attention to the plots of squared returns paired with the sample paths of market time as obtained from the estimated $\Delta g(t)$ processes. These appear in figure 5.1 of appendix 3. The three $\Delta g(t)$ processes appear quite similar, although upon closer examination, it is clear that the time deformation involving average quotes looks quite distinct from the other two specifications.

Since we have the $\Delta g(t)$ process, we may also proceed as in Müller et al. (1993) and analyze returns not in calendar time, but rather in operational time. It is a useful tool, as Müller et al. (1993) suggest to study "deseasonalized" returns. It should be noted though that while the Olsen and Associates activity scale is purely based on average (repetitive) patterns, our uses direct dynamic effects. To compute the autocorrelation function in operational time from our models, we use an approximation, namely, we define the $[\Delta \log y_t - a_1,\ \Delta \log y_{t-1} - \lambda]^2 / \Delta g(t)$ process as being normalized returns, relative to market time. Obviously, when $\Delta g(t) \equiv 1$, we recover the calendar time process. Othetwise, we recover a squared return process adjusted for serial dependence and drift which is normalized by operational time changes. This normalized process is used to compute an autocorrelation function. For comparison, we plot first the squared returns ACF in calendar time followed by the ACF computed from the Olsen and Associates time scale.[9] (See figures 5.2–5.3, appendix 3.) We observe that all operational time autocorrelation functions, namely, the Olsen and Associates and our different specifications look very different. The specification involving average spreads which were not reported in table 5.1 show significant autocorrelations at weekly, biweekly, etc. lags.

In sharp contrast, it appears from the ACF's involving operational time scales with average quotes and particularly with absolute returns, are almost white noise series which do not show any long memory properties. Judging on the basis of these ACF, it appears that the model involving absolute returns is probably the most appealing to use for the DEM/USD market.

[9]We are grateful to Michael Dacorogna for providing us with the ACF. It should be noted that the sample used in Müller et al. (1993) and the one used here is not exactly the same. We ignore this aspect here.

5.2 The JPY/USD and JPY/DEM Foreign Exchange Markets

We now turn our attention to tables 5.2 and 5.3, each covering empirical results from one market. Again, we present model specifications involving two different measures of average market activity, *nqa* and *ara*, combined with three measures of deviations introduced in the previous section.

There are several differences between the parameter estimates based on the DEM/USD sample and those reported on the JPY/USD and JPY/DEM markets. In tables 5.2 and 5.3 we notice immediately positive as well as negative signs of Θ_{ij} and we also see that some coefficients became insignificant. Exceptionally, the parameter estimates of the JPY/USD model involving the *nqa* variable (table 5.2, top panel) have similar signs as the coefficients of the analogue specification estimated from the DEM/USD data. Consequently, both models yield a similar interpretation of the volatility behavior. The operational time slows down when the number of expected quotes increases and it accelerates while the current number of quotes, the current level of spread or returns exceeds the expected values. Thus, changes in the volatility appear to be driven by the extent in which the actual market activity deviates from the average level. The results presented in the bottom panel of table 5.2 indicate that the surprise terms have the same effect in the specification involving the *ara* variable. A high level of expected returns, contrary to the average quote arrival, speeds up the operational time and the volatility adjustments.

The results based on the JPY/DEM sample are difficult to interpret. In section 2 we have pointed out several distinct distributional properties of JPY/DEM quotes. Some particular seasonal patterns of this series have also been discussed in section 4. It appears that the only variable accelerating the operational time and, hence, changes in the volatility process, is the instantaneous excess return. The coefficients on the remaining variables are positive throughout both specifications indicating an opposite effect. In conclusion, the JPY/DEM model yields results which are not plausible, and it seems appropriate to estimate volatility on this particular market within a different framework.

Table 5.1

QML Estimates of Stochastic Volatility Models with Time Deformation

20 Minute Sampling Intervals — DEM/USD Market

Model: $$\log[\Delta \log y_t - a_1 \Delta \log y_{t-1} - \lambda]^2 = -1.27 + h_t + \varsigma_t;\ h_t = [(1 - \exp(a\Delta g(t)))]b + \exp(a\Delta g(t))h_{t-1} + v_t$$

$$\Delta g(t) \approx \exp[\Theta_{qa} nqa_{t-1} + \Theta_{qd}(nqa_{t-1} - nq_{t-1}) + \Theta_{sd}(spa_{t-1} - sp_{t-1}) + \Theta_{rd}(ara_{t-1} - ar_{t-1})]$$

$$v_t \sim N(0, -\Sigma(1 - \exp(2a\Delta g(t)))/2a)$$

	(1)		(2)		(3)		(4)	
	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.
Θ_{qa}	−0.0106	0.0011	−0.0107	0.0011	−0.0114	0.0032	−0.0148	0.0016
Θ_{qd}	−0.0197	0.0029	−0.0196	0.0029	0.0320	0.0057	—	—
Θ_{sd}	−1.5040	0.0054	−1.4805	0.0053	—	—	−1.6044	0.0262
Θ_{rd}	−4.5236	0.0049	—	—	—	—	−4.7899	0.0050
a	−0.4206	0.0050	−0.3800	0.0053	−0.1357	0.0305	−0.6455	0.0226
Σ	1.1714	0.0049	1.0581	0.0049	0.3758	0.1050	1.7842	0.0114
b	−14.9361	0.0050	−14.9369	0.0050	−14.9702	0.0773	−14.8807	0.0375

$$\Delta g(t) \approx \exp[\Theta_{ra} ara_{t-1} + \Theta_{qd}(nqa_{t-1} - nq_{t-1}) + \Theta_{sd}(spa_{t-1} - sp_{t-1}) + \Theta_{rd}(ara_{t-1} - ar_{t-1})]$$

	(1)		(2)		(3)		(4)	
	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.
Θ_{ra}	−2.2965	0.0050	−2.3450	0.0050	−2.2099	0.6004	−1.8602	0.0049
Θ_{qd}	−0.0270	0.0028	−0.0270	0.0028	−0.0268	0.0047	—	—
Θ_{sd}	−1.7471	0.0161	−1.7464	0.0157	—	—	−1.8360	0.0142
Θ_{rd}	−3.0309	0.0049	—	—	—	—	−2.6463	0.0049
a	−0.7601	0.0269	−0.7418	0.0245	−0.1257	0.0189	−0.8759	0.0276
Σ	2.0766	0.0115	2.0260	0.0106	0.3382	0.0629	2.2712	0.0120
b	−14.8290	0.0060	−14.8193	0.0119	−14.8360	0.0614	−14.6385	0.0106

Table 5.2

QML Estimates of Stochastic Volatility Models with Time Deformation

20 Minute Sampling Intervals — JPY/USD Market

Model: $\log[\Delta \log y_t - a_1 \Delta \log y_{t-1} - \lambda]^2 = -1.27 + h_t + \varsigma_t;\ h_t = [(1-\exp(a\Delta g(t)))]b + \exp(a\Delta g(t))h_{t-1} + v_t$

$$\Delta g(t) \approx \exp[\Theta_{qa} nqa_{t-1} + \Theta_{qd}(nqa_{t-1} - nq_{t-1}) + \Theta_{sd}(spa_{t-1} - sp_{t-1}) + \Theta_{rd}(ara_{t-1} - ar_{t-1})]$$

$$v_t \sim N(0, -\Sigma(1 - \exp(2a\Delta g(t)))/2a)$$

	(1)		(2)		(3)		(4)	
	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.
Θ_{qa}	−0.0153	0.0043	−0.0167	0.0066	−0.0070	0.0024	−0.0070	0.0042
Θ_{qd}	−0.0239	0.0036	−0.0243	0.0083	—	—	—	—
Θ_{sd}	−0.2002	0.0049	—	—	0.2970	0.1401	—	—
Θ_{rd}	−0.8204	0.0049	—	—	—	—	−0.5879	0.0049
a	−0.2189	0.0053	−0.2152	0.0411	−0.1943	0.0069	−0.1983	0.0399
Σ	0.6899	0.0111	0.6777	0.1756	0.5801	0.0239	0.5951	0.0079
b	−14.9240	0.0050	−14.9306	0.0596	−14.7899	0.0212	−14.7950	0.0050

$$\Delta g(t) \approx \exp[\Theta_{ra} ara_{t-1} + \Theta_{qd}(nqa_{t-1} - nq_{t-1}) + \Theta_{sd}(spa_{t-1} - sp_{t-1}) + \Theta_{rd}(ara_{t-1} - ar_{t-1})]$$

	(1)		(2)		(3)		(4)	
	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.
Θ_{ra}	6.2526	0.0049	6.2884	2.6309	6.0877	3.8865	5.2324	0.0049
Θ_{qd}	−0.0178	0.0041	−0.0181	0.0054	—	—	—	—
Θ_{sd}	−0.2864	0.0050	—	—	−0.3251	0.1398	—	—
Θ_{rd}	−0.7404	0.0049	—	—	—	—	−0.5565	0.0049
a	−0.2063	0.0161	−0.2063	0.0202	−0.1921	0.0068	−0.1943	0.0154
Σ	0.6357	0.0643	0.6354	0.0804	0.5702	0.0234	0.5777	0.0596
b	−14.8784	0.0339	−14.8860	0.0388	−14.7786	0.0213	−14.7838	0.0024

Table 5.3

QML Estimates of Stochastic Volatility Models with Time Deformation

20 Minute Sampling Intervals — JPY/DEM Market

Model: $\log[\Delta \log y_t - a_1 \Delta \log y_{t-1} - \lambda]^2 = -1.27 + h_t + \varsigma_t;\ h_t = [(1-\exp(a\Delta g(t)))]b + \exp(a\Delta g(t))h_{t-1} + v_t$

$$\Delta g(t) \approx \exp[\Theta_{qa} nqa_{t-1} + \Theta_{qd}(nqa_{t-1} - nq_{t-1}) + \Theta_{sd}(spa_{t-1} - sp_{t-1}) + \Theta_{rd}(ara_{t-1} - ar_{t-1})]$$

$$v_t \sim N(0, -\Sigma(1-\exp(2a\Delta g(t)))/2a)$$

	(1)		(2)		(3)		(4)	
	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.
Θ_{qa}	0.0273	0.0074	0.0022	0.0132	0.0186	0.0134	−0.0050	0.0131
Θ_{qd}	−0.0204	0.0097	0.0201	0.0111	—	—	—	—
Θ_{sd}	0.4924	0.0054	—	—	0.4822	0.0800	—	—
Θ_{rd}	−0.2240	0.0052	—	—	—	—	0.1387	0.5881
a	−0.1773	0.0098	−0.1743	0.0088	−0.1781	0.0091	−0.1767	0.0089
Σ	0.3195	0.0272	0.3163	0.0212	0.3232	0.0215	0.3241	0.0216
b	−14.3923	0.0022	−14.3674	0.0202	−14.4061	0.0199	−14.3817	0.0196

$$\Delta g(t) \approx \exp[\Theta_{ra} ara_{t-1} + \Theta_{qd}(nqa_{t-1} - nq_{t-1}) + \Theta_{sd}(spa_{t-1} - sp_{t-1}) + \Theta_{rd}(ara_{t-1} - ar_{t-1})]$$

	(1)		(2)		(3)		(4)	
	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.	Est.	St. Er.
Θ_{ra}	6.6248	5.4780	7.7326	5.6690	6.4390	3.0797	7.6165	0.0052
Θ_{qd}	−0.0214	0.0116	0.0195	0.0111	—	—	—	—
Θ_{sd}	−0.4577	0.0787	—	—	−0.4573	0.0978	—	—
Θ_{rd}	−0.2292	0.6664	—	—	—	—	−0.1189	0.0054
a	−0.1805	0.0093	−0.1748	0.0088	−0.1812	0.0171	−0.1750	0.0147
Σ	0.3276	0.0220	0.3178	0.0212	0.3309	0.0417	0.3199	0.0356
b	−14.4100	0.0198	−14.3696	0.0196	−14.4160	0.0252	−14.3790	0.0133

6. CONCLUSIONS

In this paper we discussed the dynamics of three exchange markets: DEM/USD, JPY/USD and JPY/DEM, and we proposed a stochastic volatility model for exchange rates sampled at high frequencies.

We first examined the complexity of market dynamics emphasizing the seasonal patterns in return, bid, and ask. The analysis has been based both on unequally spaced data as well as on series sampled at fixed 20 minute intervals. We have pointed out that the choice of the time scale is crucial for the accuracy and the informational content of the results. In the tick-by-tick records, we observed some interesting shifts in the *entire* distributions from one month to the other and even throughout the week. The equally spaced data exhibit similar radical changes in the behavior of the empirical distributions through time. The complexity of the seasonals in high frequency records requires, thus, a more sophisticated framework than simple mean shift models of standard adjustment techniques developed in the (macro) time series analysis. Finally, we presented evidence that the usual geometric average of bids and asks is an appropriate measure of returns on the 20 minute time scale but is an unreliable indicator of mean price changes in the tick-by-tick records.

Next, we investigated a new approach to deal with the seasonal effects in high frequency data and proposed a time deformation framework of stochastic volatility. It is worth emphasizing that it is the first attempt to fit this type of model to high frequency exchange rate series. We examined two specifications of the relationship between the volatility of quotes and the expected values of some relevant variables approximating the market activity as well as the instantaneous deviations from their average behavior. In general, the models successfully explained the market dynamics at least in two out of three data sets.

Appendix 1

Figure 2.1

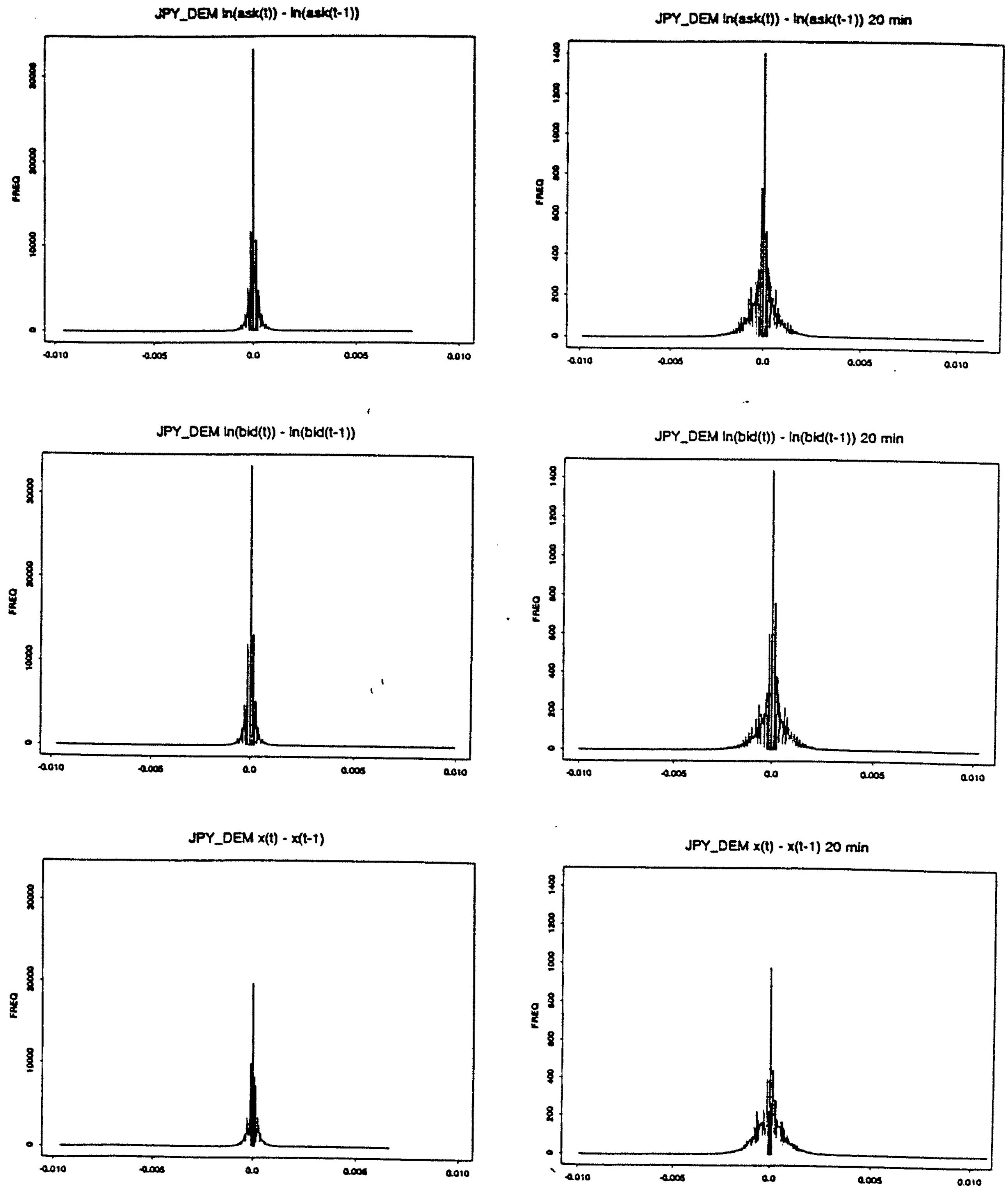


Figure 2.2

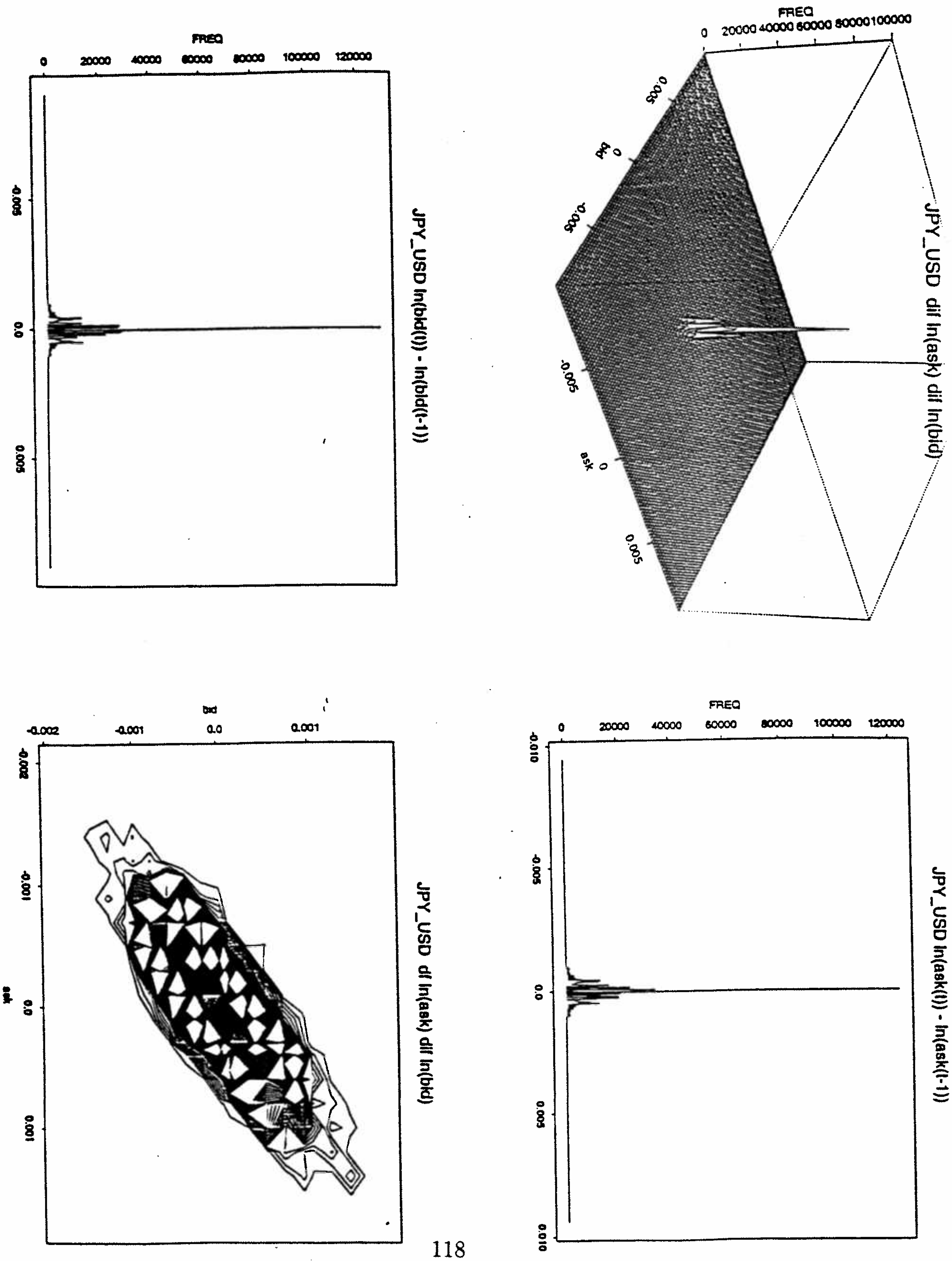

JPY_USD dif ln(ask) dif ln(bid)
JPY_USD ln(ask(t)) - ln(ask(t-1))
JPY_USD ln(bid(t)) - ln(bid(t-1))
JPY_USD dif ln(ask) dif ln(bid)
FREQ

Figure 2.3

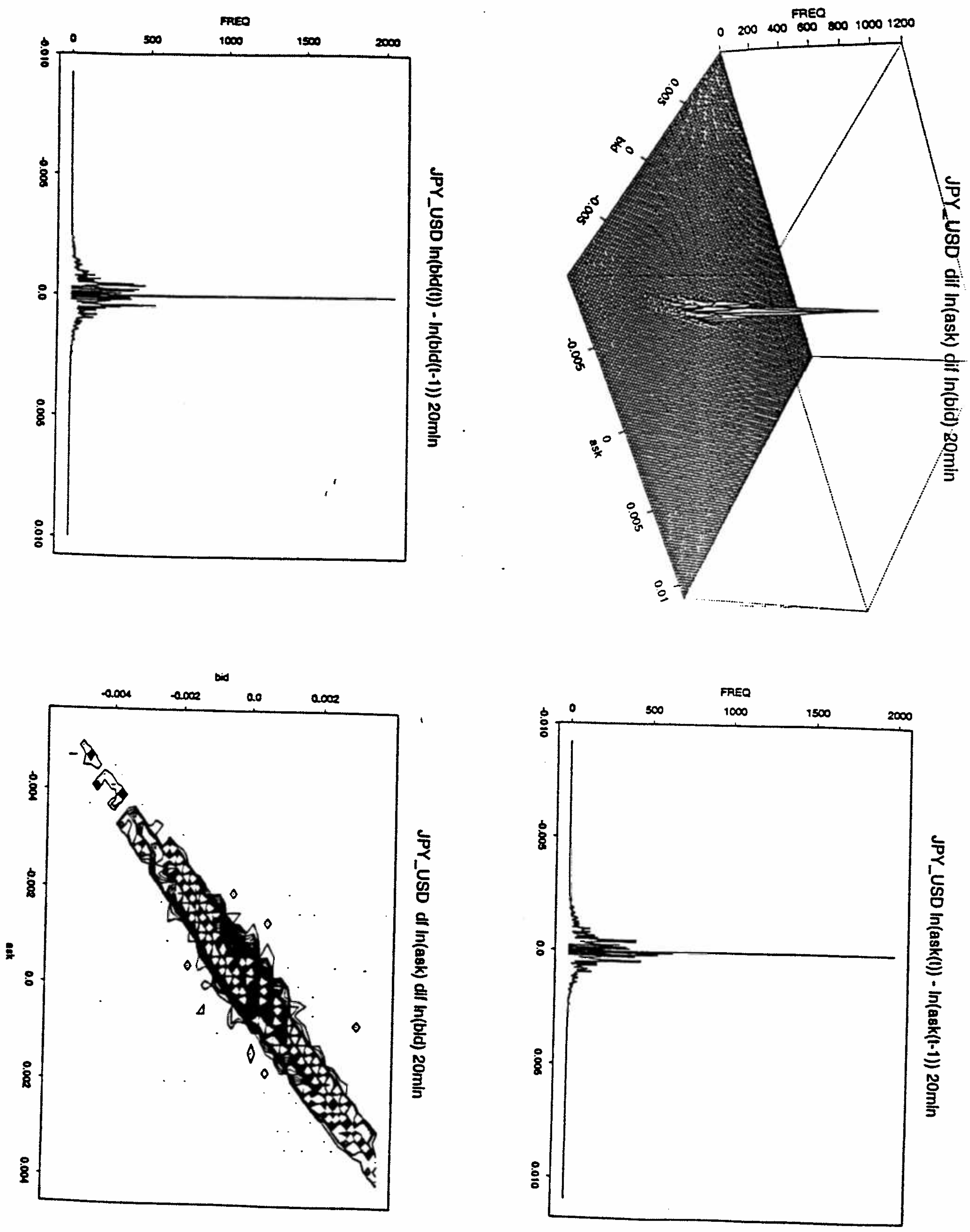

JPY_USD dif ln(ask) dif ln(bid) 20min
FREQ
0 200 400 600 800 1000 1200
bid
ask
-0.005
0
0.005
0.01
JPY_USD ln(bid(t)) - ln(bid(t-1)) 20min
FREQ
0 500 1000 1500 2000
-0.010 -0.005 0.0 0.005 0.010
JPY_USD ln(ask(t)) - ln(ask(t-1)) 20min
JPY_USD dif ln(ask) dif ln(bid) 20min
-0.004 -0.002 0.0 0.002 0.004

Figure 2.4

Bivariate Monthly Histograms of JPY/USD Quotes (Real Time)

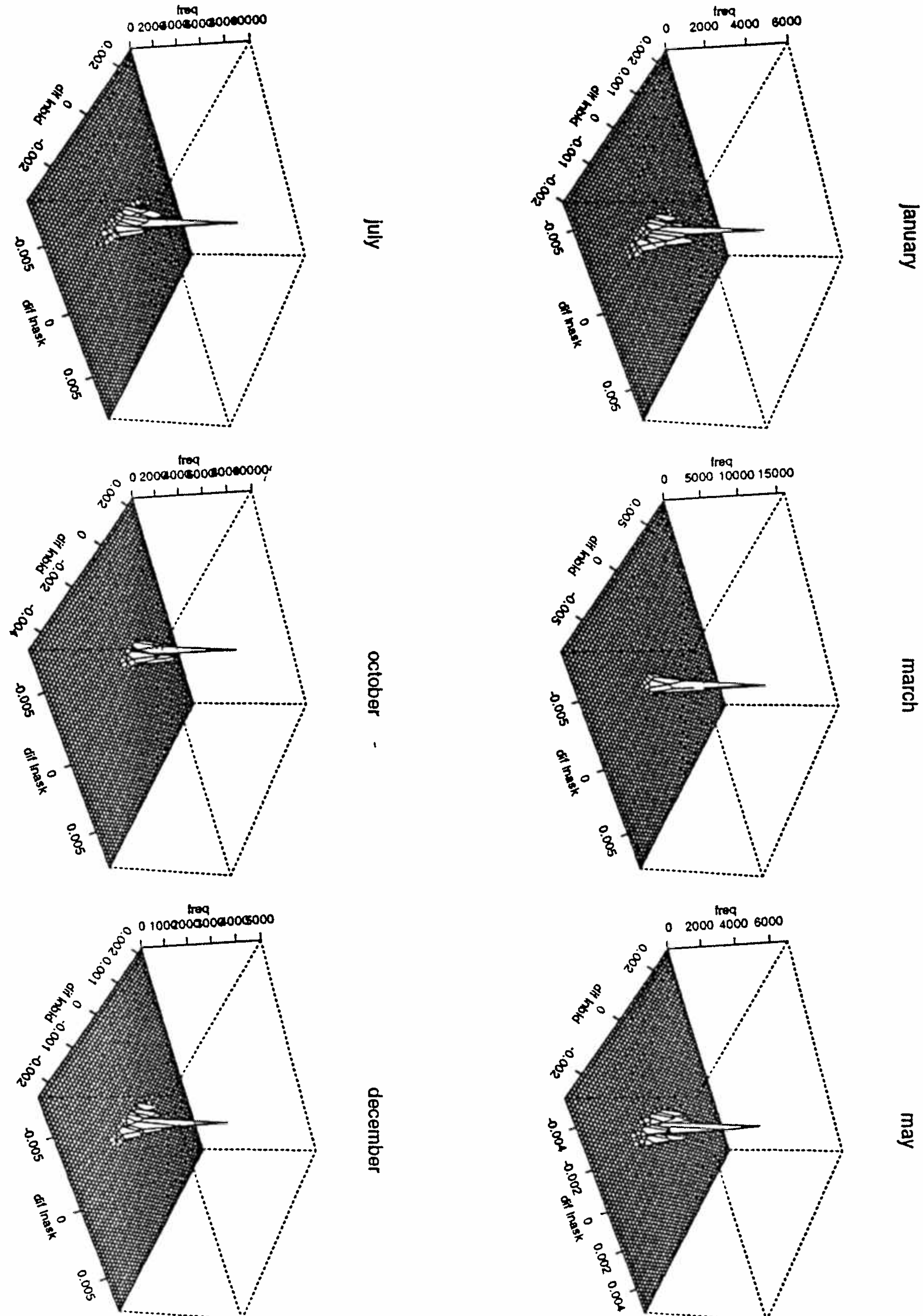


Figure 2.5

Bivariate Monthly Histograms of JPY/USD Quotes (20 Minutes)

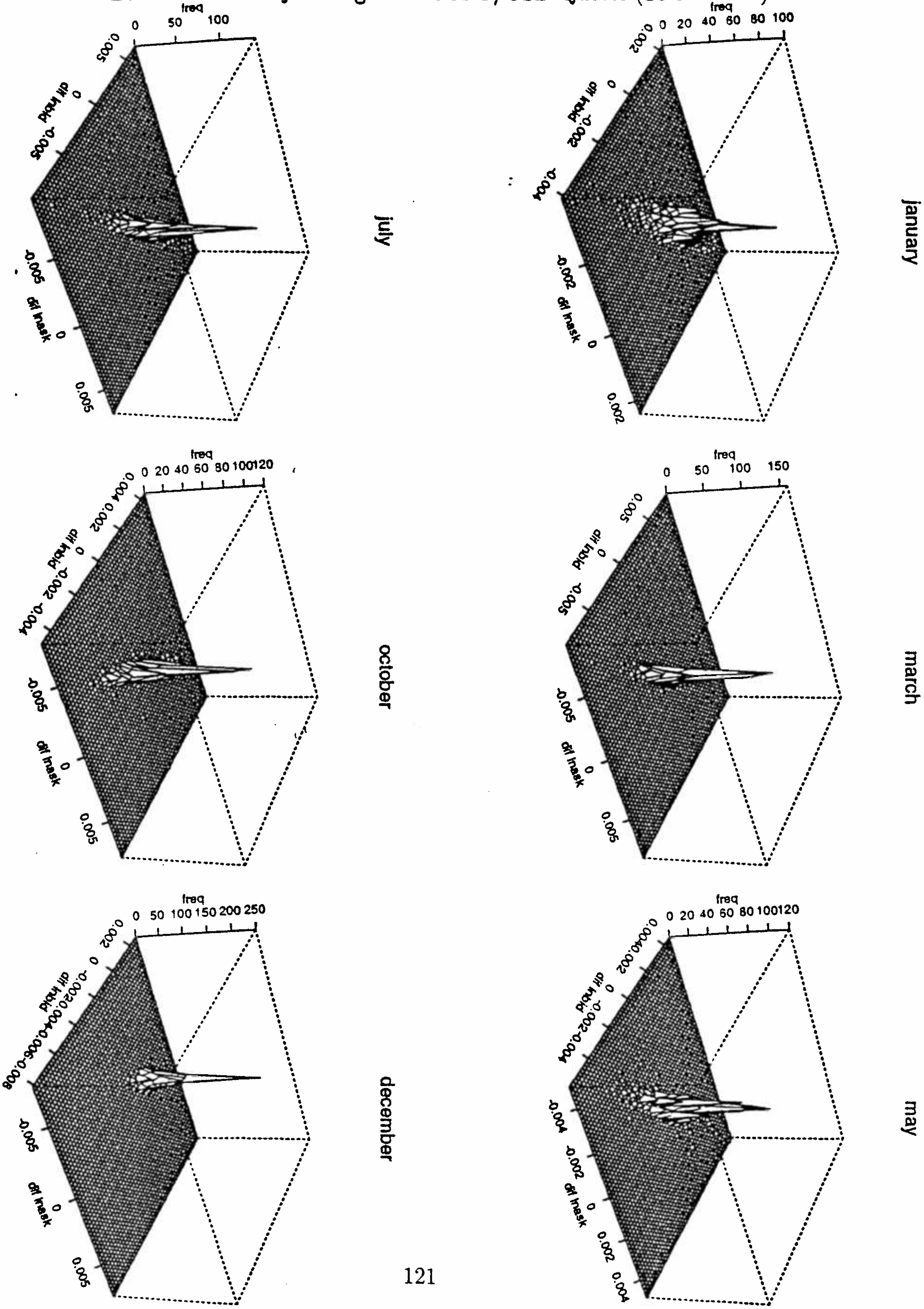


Figure 2.6

Bivariate Daily Histograms of DEM/USD Quotes (Real Time)

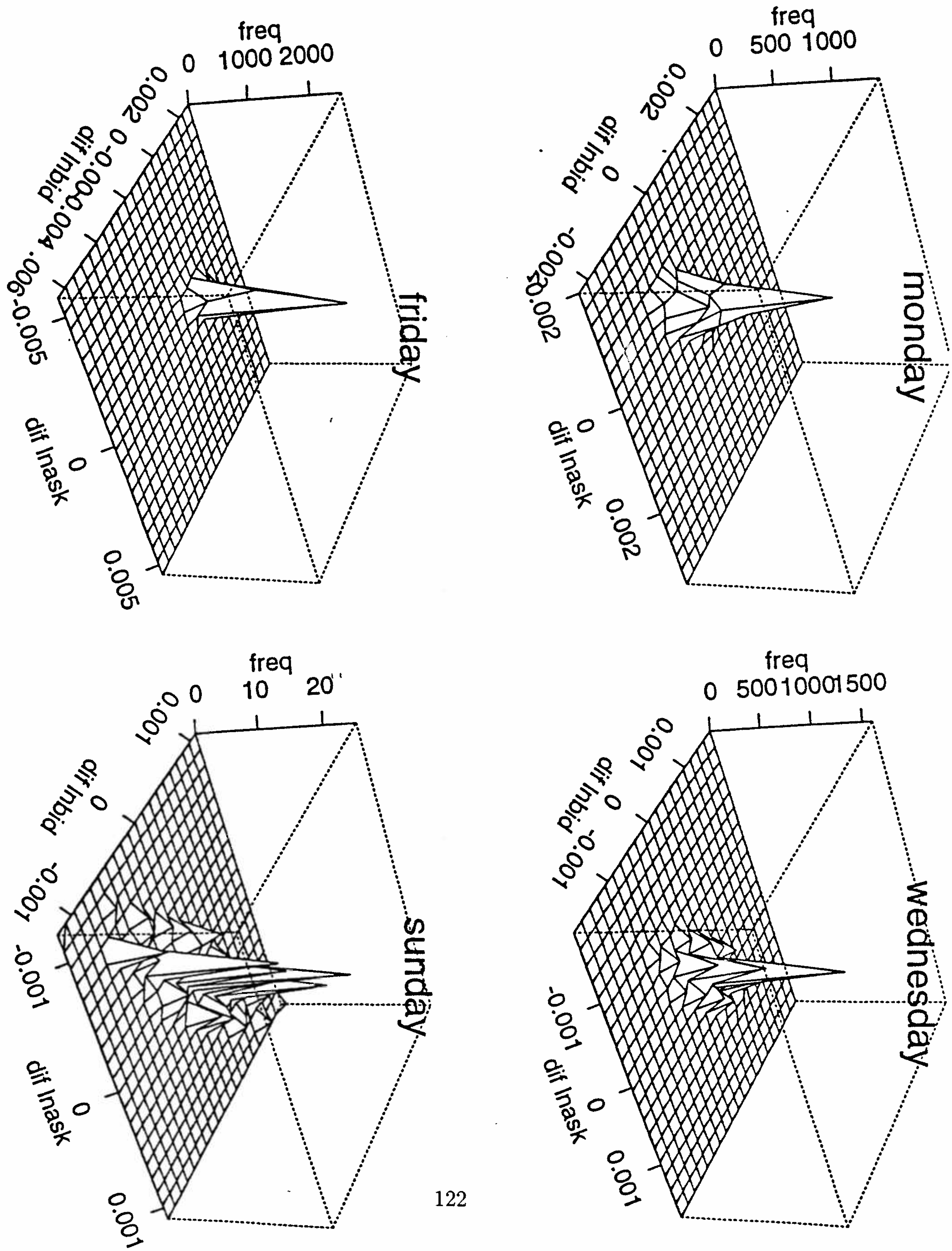


Figure 2.7

Bivariate Daily Histograms of JPY/USD Quotes (20 Minutes)

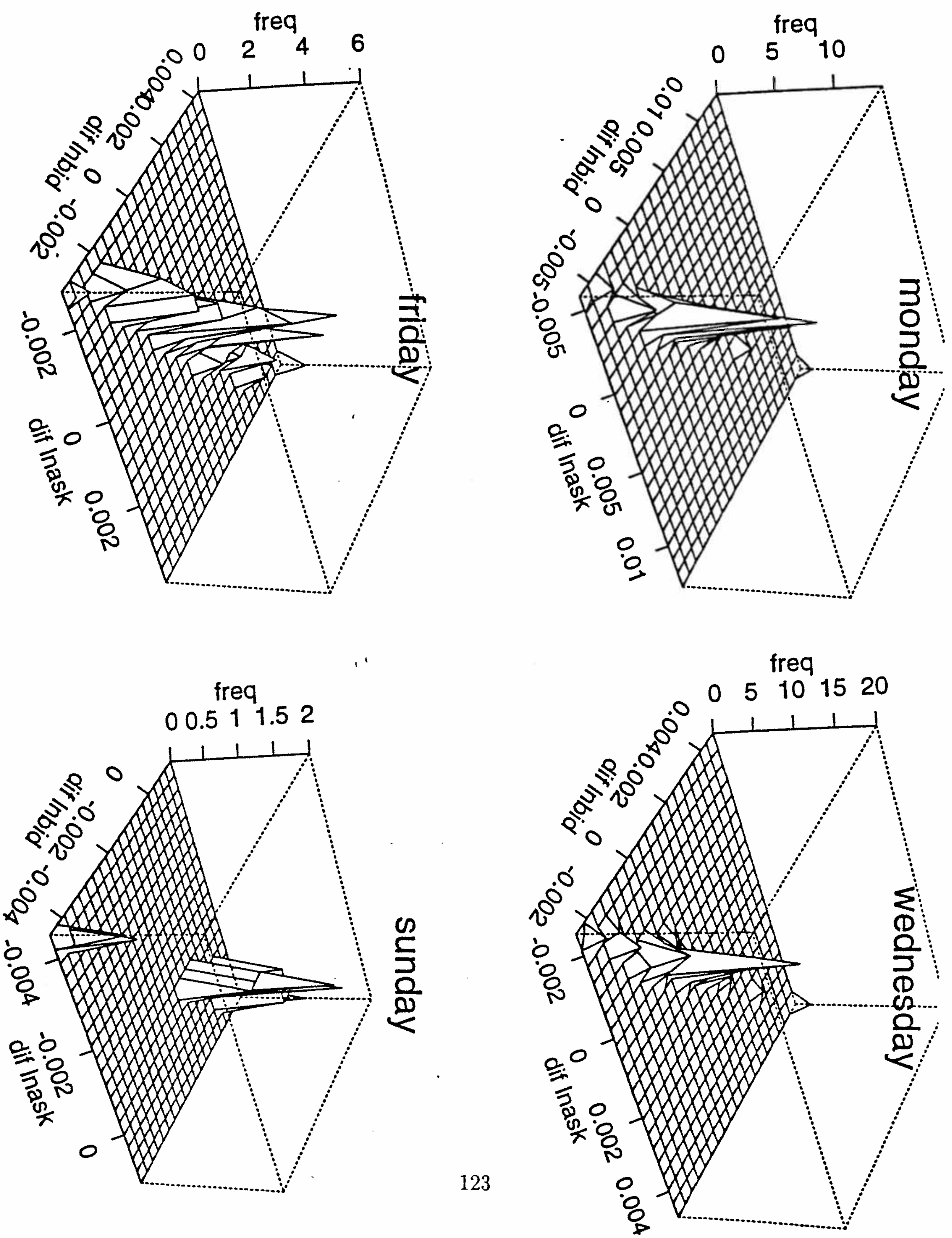


Figure 2.8

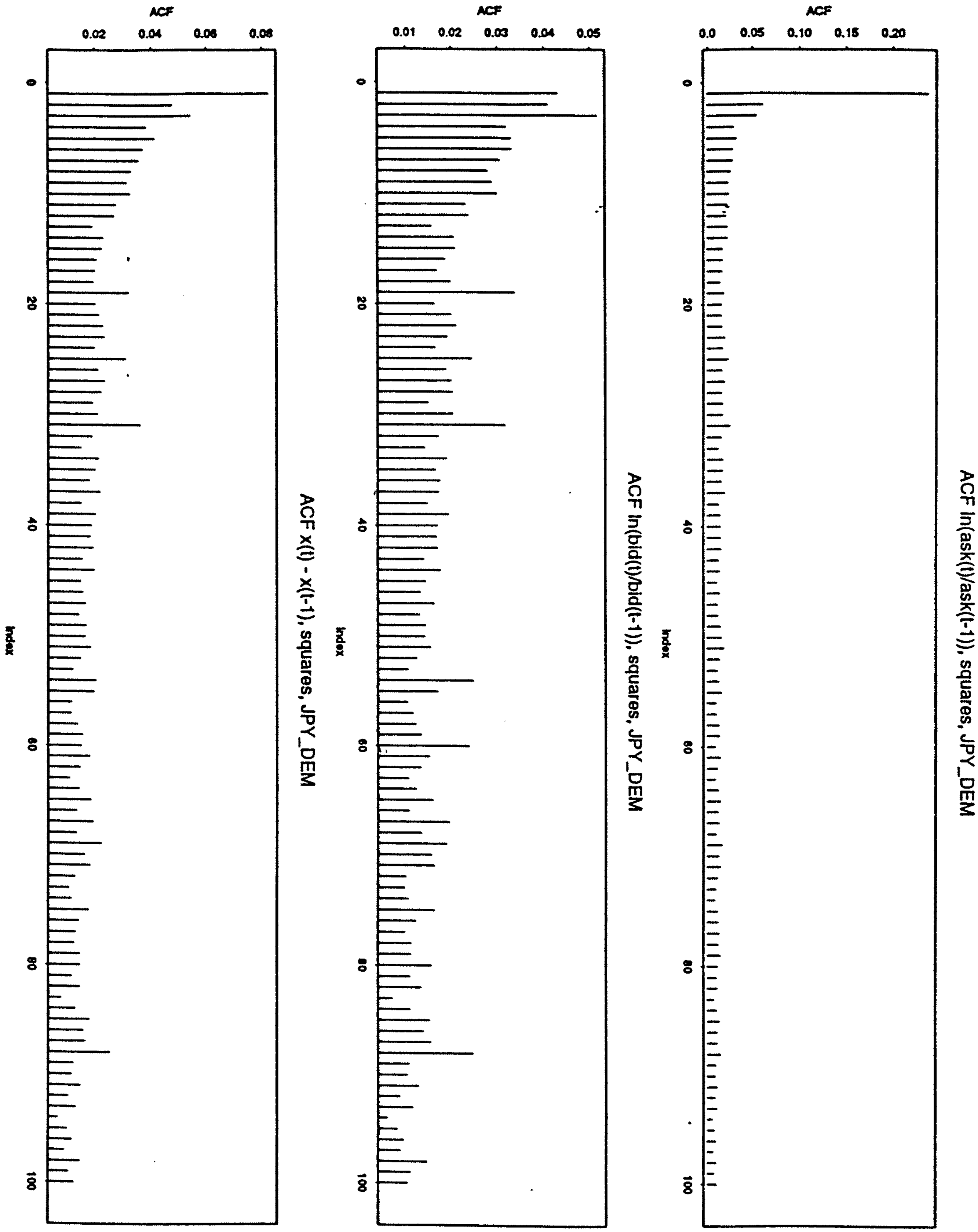
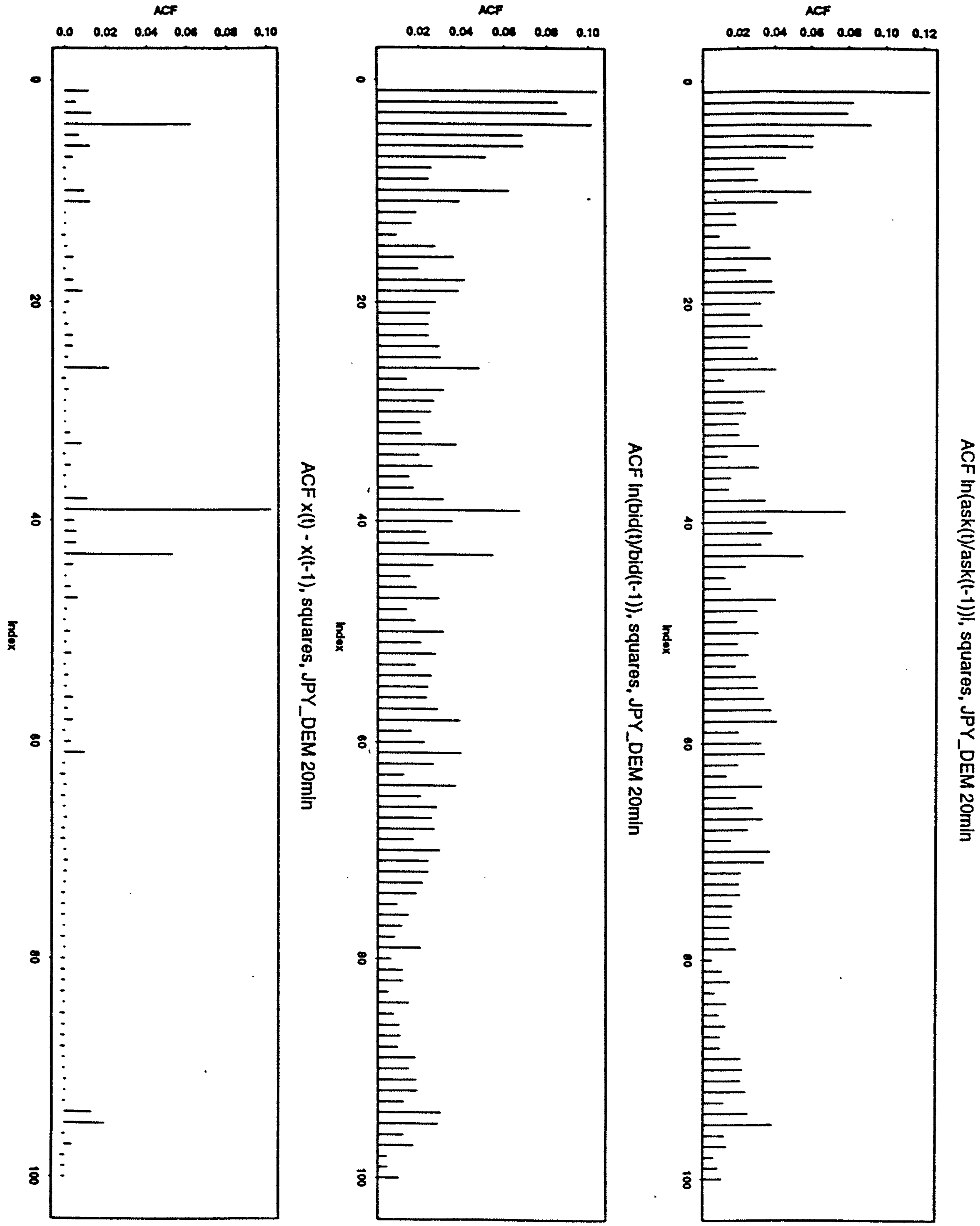

ACF ln(ask(t)/ask(t-1)), squares, JPY_DEM
ACF ln(bid(t)/bid(t-1)), squares, JPY_DEM
ACF x(t) - x(t-1), squares, JPY_DEM
ACF
Index
0.0
0.05
0.10
0.15
0.20
0.01
0.02
0.03
0.04
0.05
0.02
0.04
0.06
0.08
0
20
40
60
80
100

Figure 2.9


ACF x(t) - x(t-1), squares, JPY_DEM 20min
ACF ln(bid(t)/bid(t-1)), squares, JPY_DEM 20min
ACF ln(ask(t)/ask(t-1))l, squares, JPY_DEM 20min
ACF
Index

Appendix 2

Figure 4.1

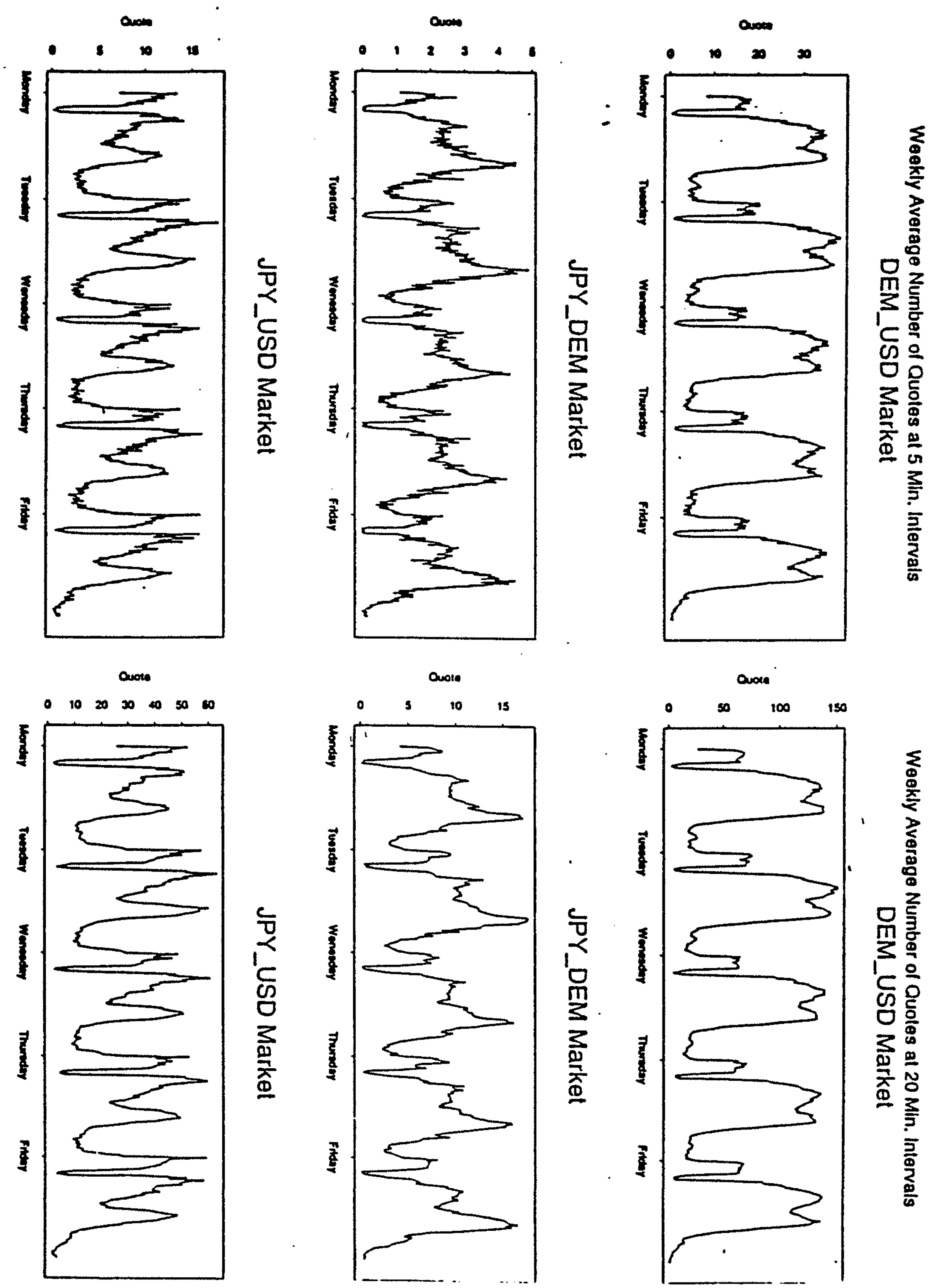


Figure 4.2

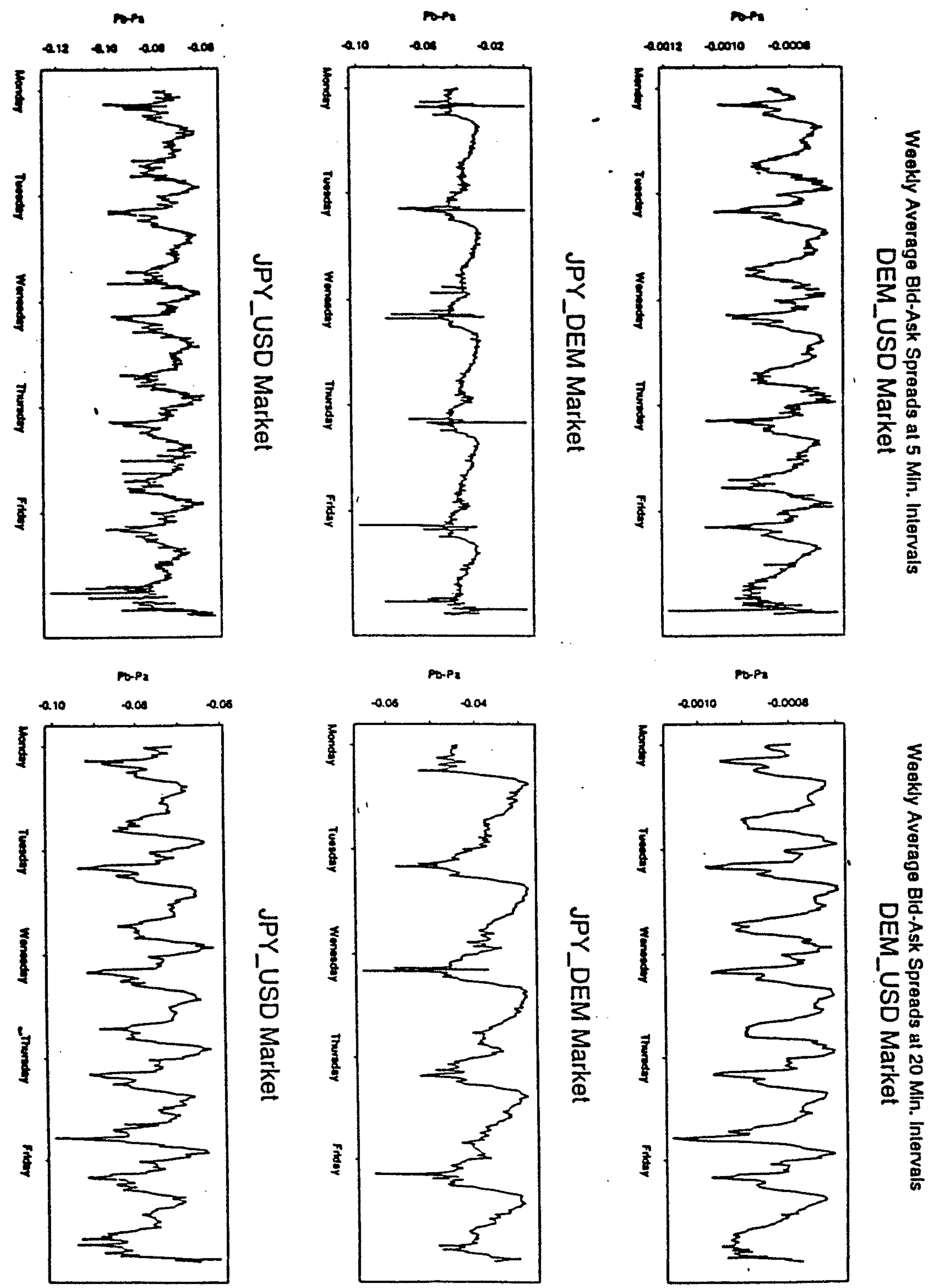

Weekly Average Bid-Ask Spreads at 5 Min. Intervals
DEM_USD Market
JPY_DEM Market
JPY_USD Market
Weekly Average Bid-Ask Spreads at 20 Min. Intervals
DEM_USD Market
JPY_DEM Market
JPY_USD Market
Pb-Pa
Monday
Tuesday
Thursday
Friday

Figure 4.3

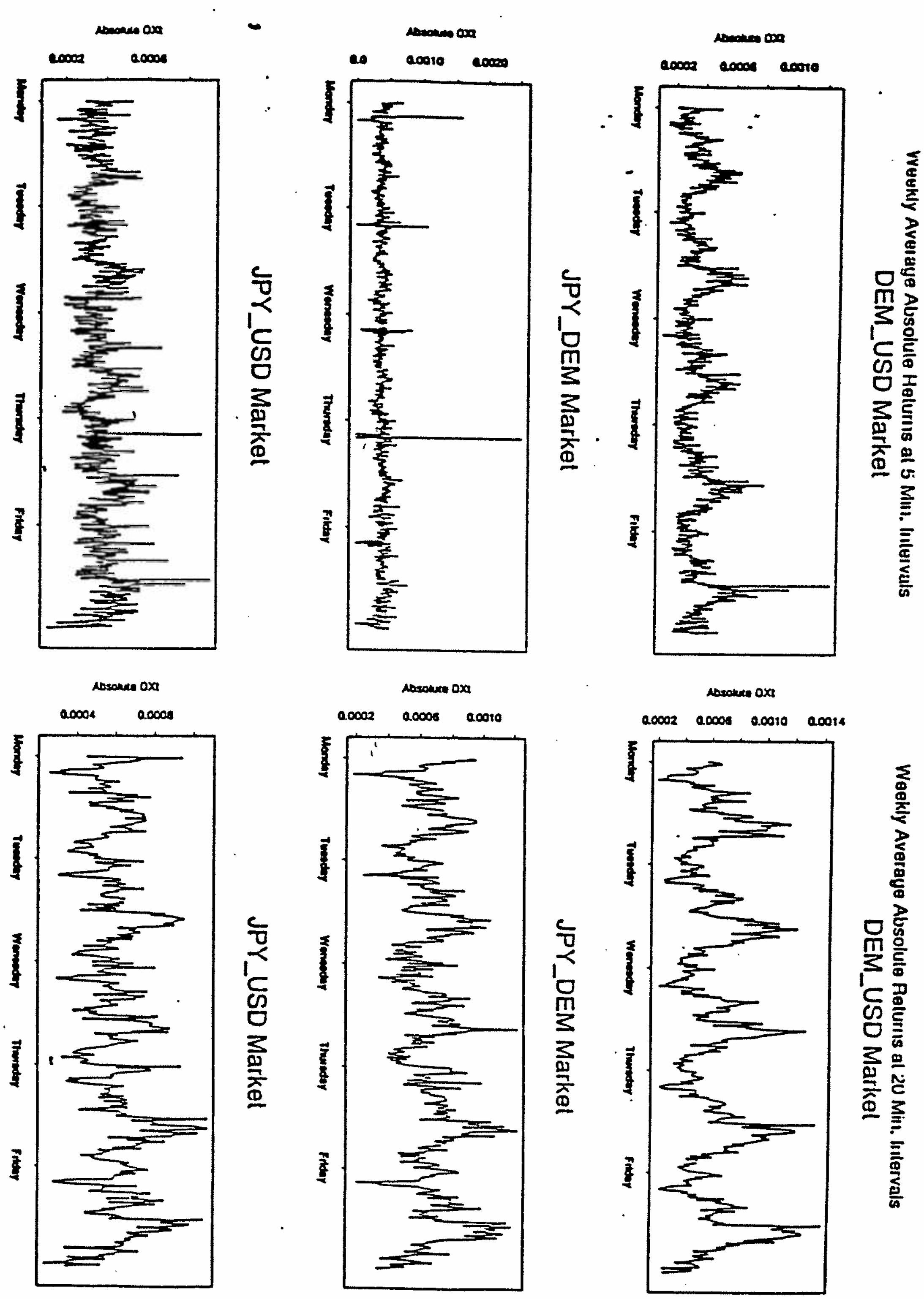

Weekly Average Absolute Returns at 5 Min. Intervals
DEM_USD Market
JPY_DEM Market
JPY_USD Market
Weekly Average Absolute Returns at 20 Min. Intervals
DEM_USD Market
JPY_DEM Market
JPY_USD Market

Figure 4.4

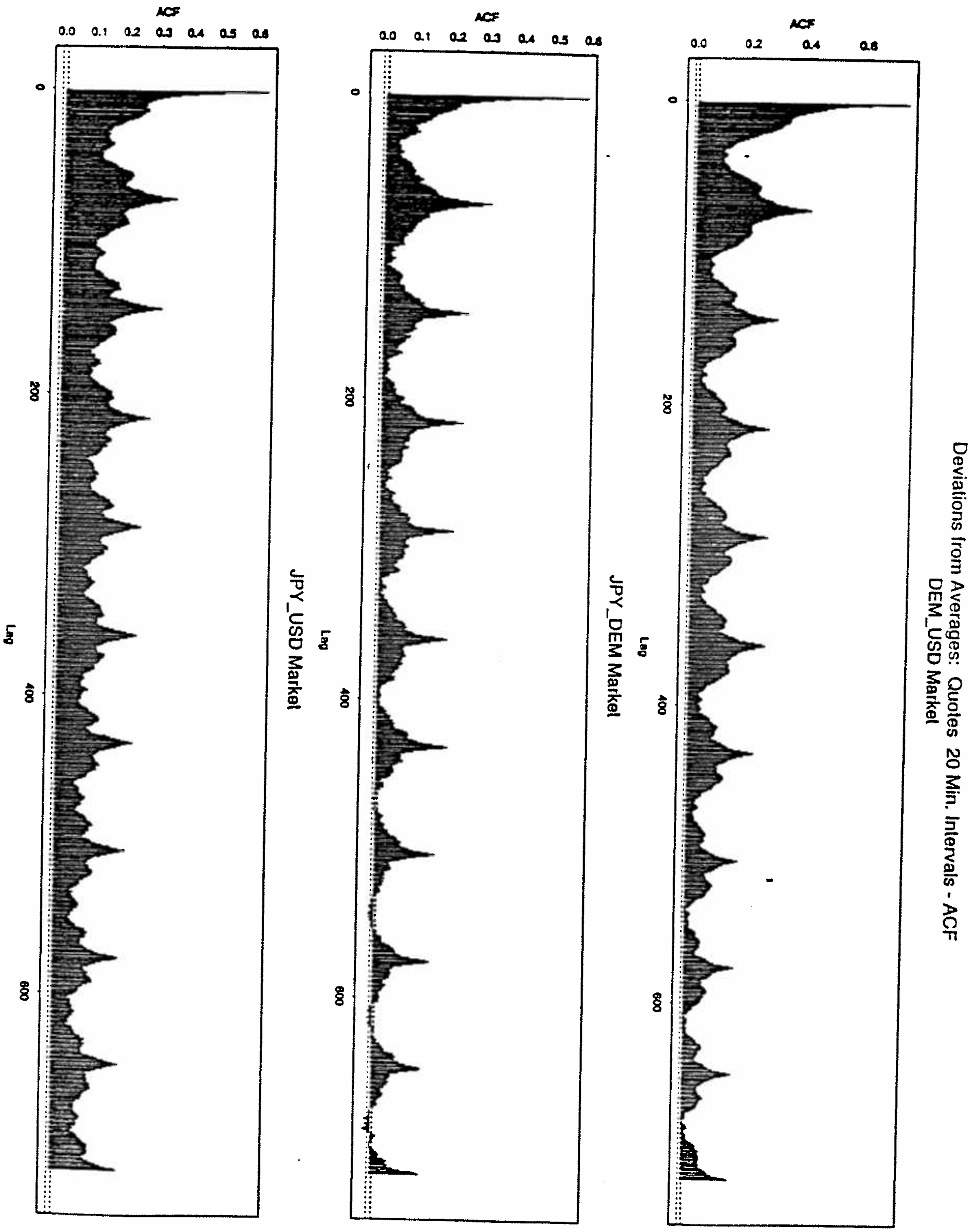

Deviations from Averages: Quotes 20 Min. Intervals - ACF
DEM_USD Market
JPY_DEM Market
JPY_USD Market
ACF
Lag

Figure 4.5

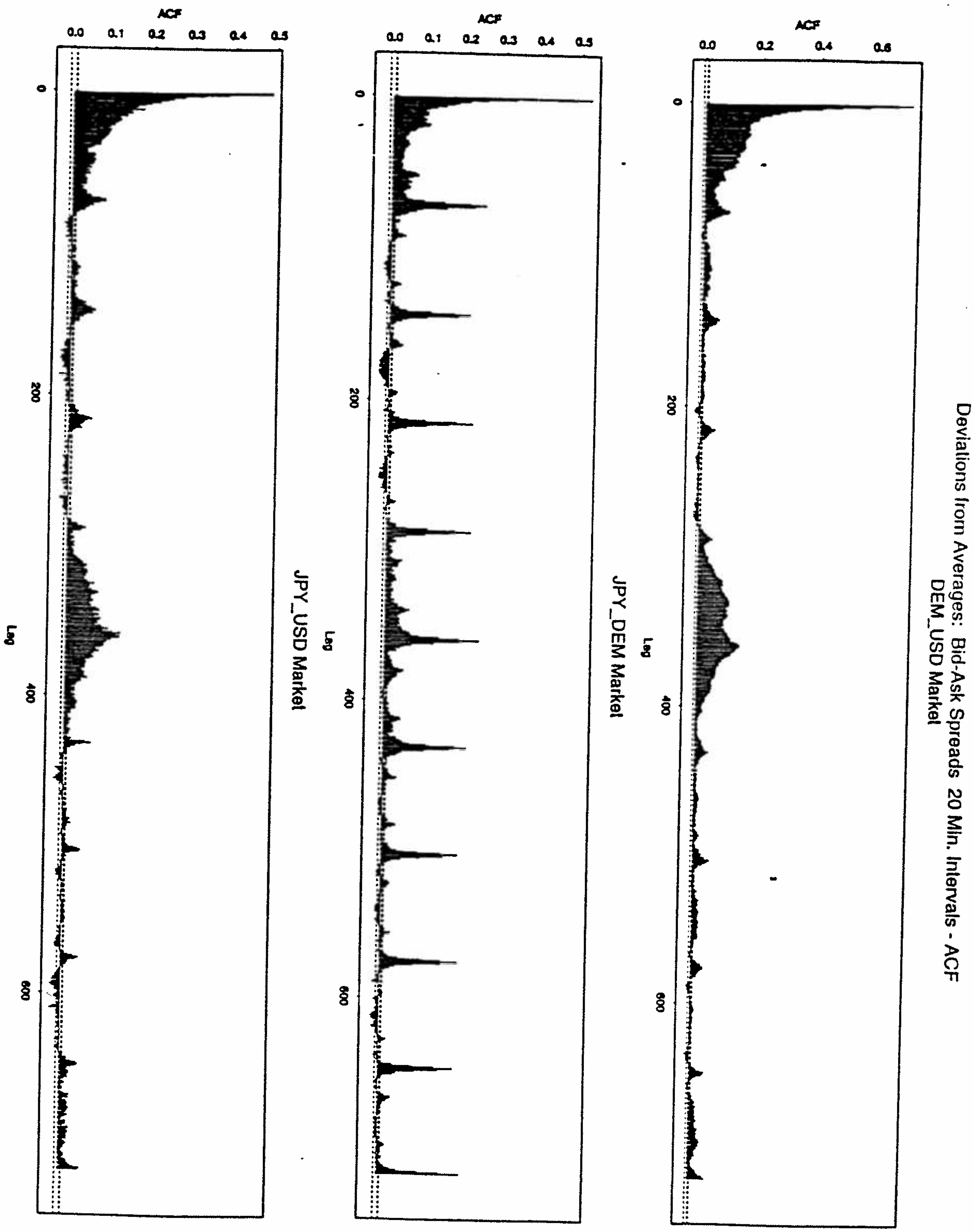
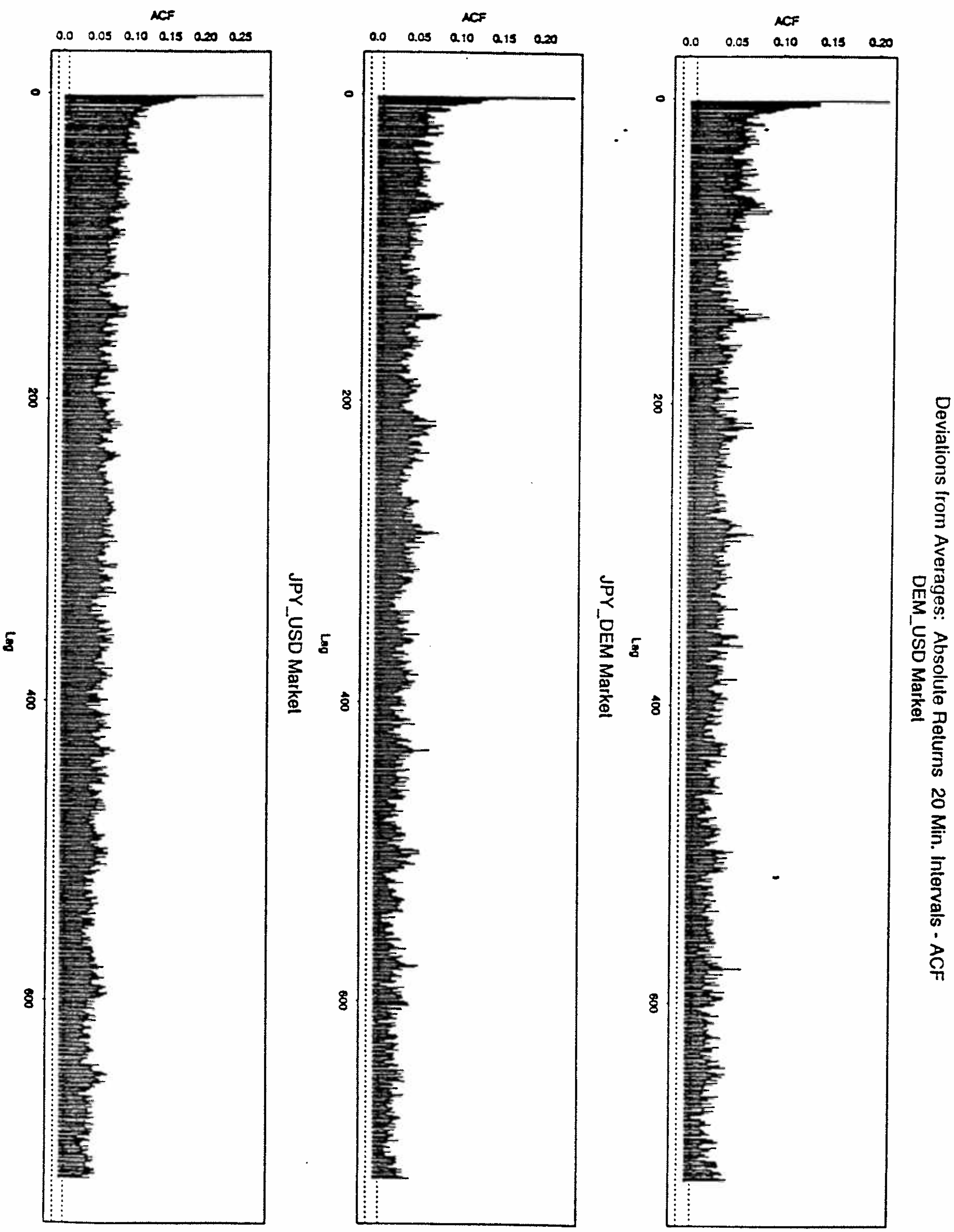

Deviations from Averages: Bid-Ask Spreads 20 Min. Intervals - ACF
DEM_USD Market
JPY_DEM Market
JPY_USD Market
ACF
Lag
0.0
0.2
0.4
0.6
0.1
0.3
0.5
0
200
400
600

Figure 4.6


Deviations from Averages: Absolute Returns 20 Min. Intervals - ACF
DEM_USD Market
JPY_DEM Market
JPY_USD Market
ACF
Lag

Appendix 3

Figure 5.1

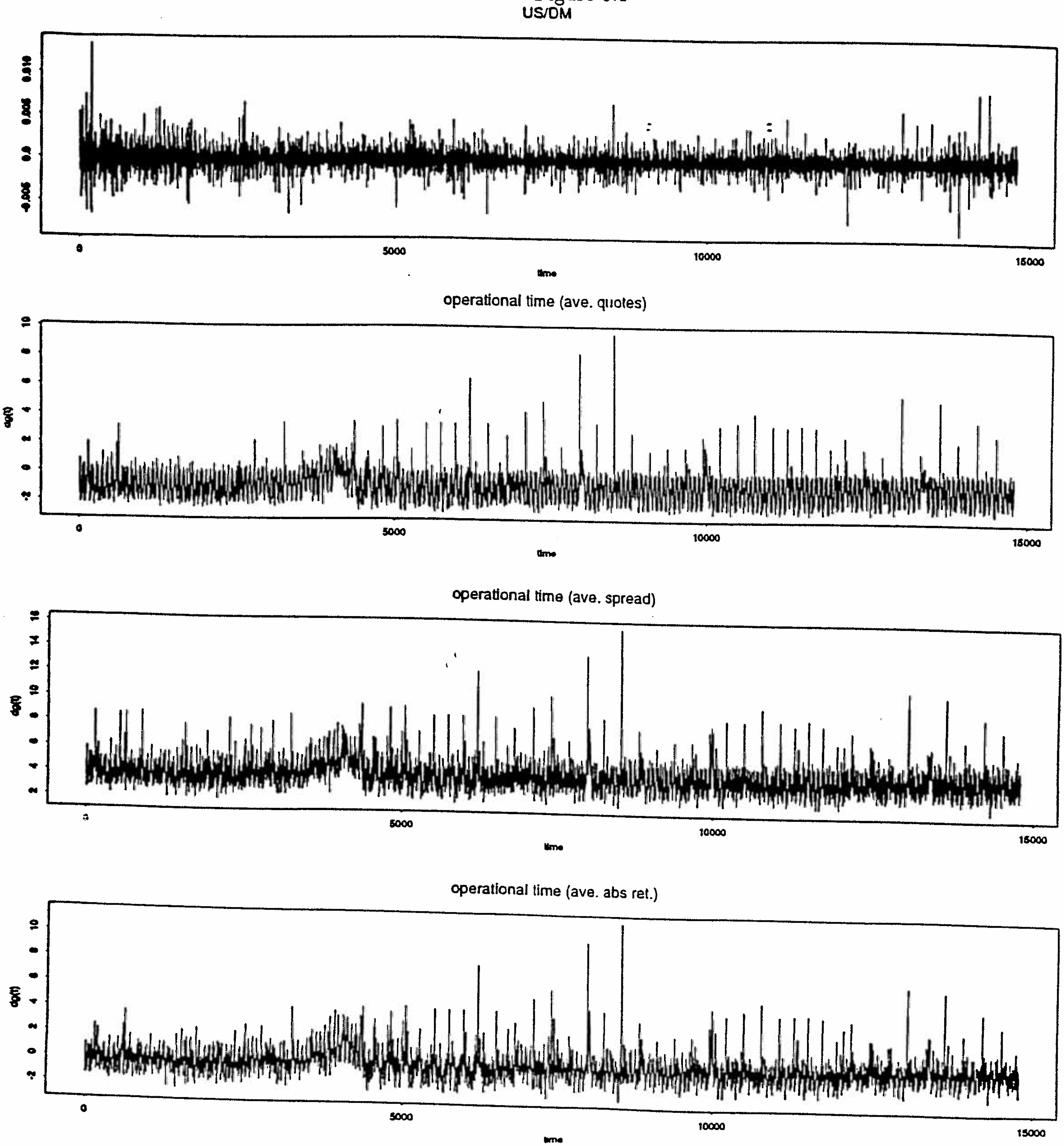
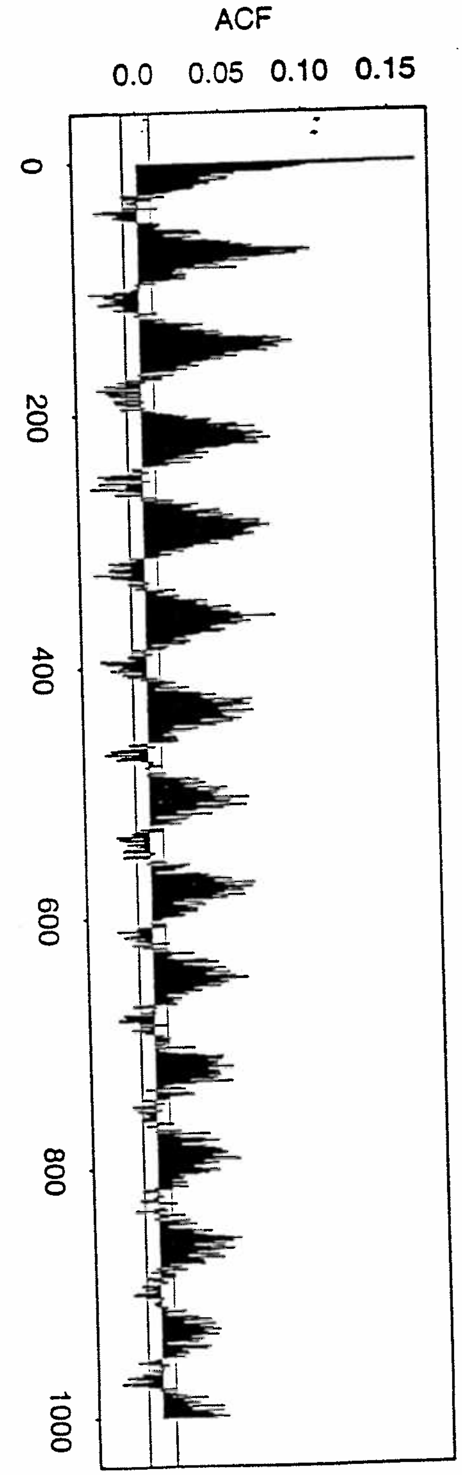
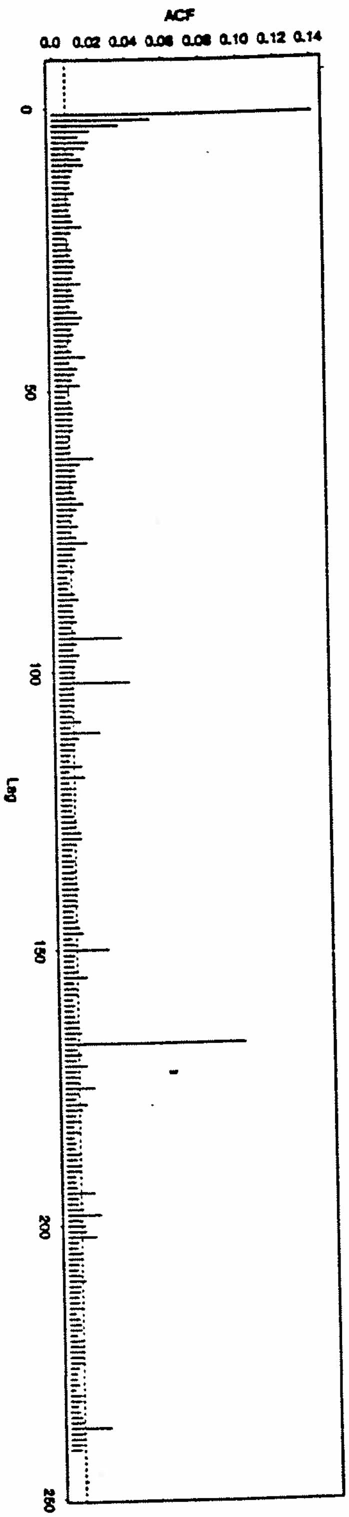


Figure 5.2

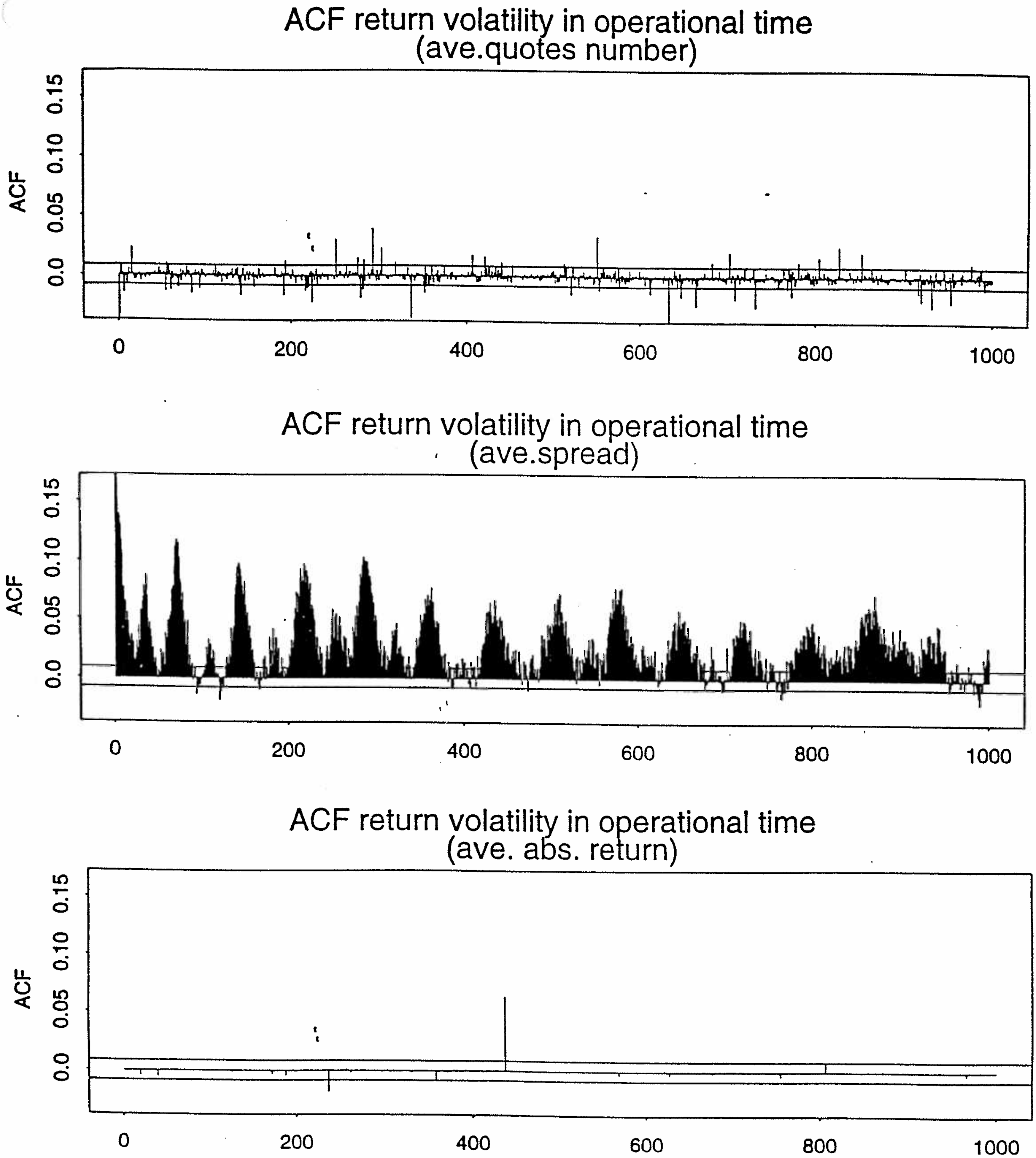
ACF return volatility





ACF for squared returns over twenty business minutes - O&A activity scale
DEM_USD Market

Figure 5.3


ACF return volatility in operational time
(ave.quotes number)
ACF
0.0 0.05 0.10 0.15
0 200 400 600 800 1000
ACF return volatility in operational time
(ave.spread)
ACF
0.0 0.05 0.10 0.15
0 200 400 600 800 1000
ACF return volatility in operational time
(ave. abs. return)
ACF
0.0 0.05 0.10 0.15
0 200 400 600 800 1000

REFERENCES

Anderson, T. and T. Bollerslev (1994), "Intraday Seasonality and Volatility Persistence in Financial Markets," Discussion Paper, Nothwestern University.

Baillie, R.T. and T. Bollerslev (1991), "Intra-Day and Inter-Market Volatility in Foreign Exchange Rates," *Review of Economic Studies* 58, 565–585.

Bessembinder, H. and P.J. Seguin (1993), "Price Volatility, Trading Volume and Market Depth: Evidence from Futures Markets," *Journal of Financial and Quantitative Analysis* 28(1), 21–39.

Bollerslev, T. and I. Domowitz (1993), "Trading Patterns and Prices in the Interbank Foreign Exchange Market," *The Journal of Finance* Vol. XLVIII, No. 4, pp. 1421–1443.

Bollerslev, T. and E. Ghysels (1994), "Periodic Autoregressive Conditional Heteroskedasticity," *Journal of Business and Economic Statistics* (forthcoming).

Chesney, M. and L. Scott (1989), "Pricing European Currency Option: A Comparison of the Modified Black-Scholes Model and a Random Variance Model," *Journal of Financial and Quantitative Analysis* 24, 267–284.

Clark, P.K. (1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica* 41(1), 135–156.

Dacorogna, M.M., C.L. Gauvreau, U.A. Müller, R.B. Olsen and O.V. Pictet (1992), "Short Term Forecasting Models of Foreign Exchange Rates," Presentation at the IBM Summer Research Institute in Oberlech, Austria on "Advanced Applications in Finance, Investment and Banking," July 27 to 31, 1992.

Dacorogna, M.M., U.A. Müller, R.J. Nagler, R.B. Olsen and O.V. Pictet (1993), "A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market," *Journal of International Money and Finance* 12, 413–438.

Drost, F.C. and T.E. Nijman (1993), "Temporal Aggregation of GARCH Processes," *Econometrica* 61, pp. 909–998.

Easley, D. and M. O'Hara (1992), "Time and the Process of Security Price Adjustment," *Journal of Finance* XLVII(2), 577–605.

Feinstone, L.J. (1987), "Minute by Minute: Efficiency, Normality and Randomness in Intra-Daily Asset Prices," *Journal of Applied Econometrics* 2, 193–214.

French, K. (1980), "Stock Returns and the Weekend Effect," *Journal of Financial Economics* 8, 55–69.

French, K.R., G.W. Schwert and R.F. Stambaugh (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics* 19, 3–29.

Gallant, A.R., P.E. Rossi, and G. Tauchen (1992), "Stock Prices and Volume," *Review of Financial Studies* 5, 871–908.

Gallant, A.R. and G. Tauchen (1994), "Which Moments to Match?," *Econometric Theory* (forthcoming).

Ghysels, E., C. Gouriéroux and J. Jasiak (1994), "Market Time and Asset Price Movements: Theory and Estimation," Discussion Paper CIRANO and CREST.

Ghysels, E. and J. Jasiak (1994), "Stochastic Volatility and Time Deformation: An Application to Trading Volume and Leverage Effects," Working Paper C.R.D.E.

Ghysels, E. (1994), "On the Economics and Econometrics of Seasonality," in C.A. Sins (ed.), *Advances in Econometrics* — 6th World Congress, Cambridge University Press.

Goffe, W.L., Ferrier, G.D. and J. Rogers (1994), "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics* 60, 65–100.

Goodhart, C.A.E., and L. Figliuoli (1988), "The Geographical Location of the Foreign Exchange Market: A Test of an 'Island' Hypothesis," LSE financial markets group, Discussion Paper No. 38.

Goodhart, C.A.E., and L. Figliuoli (1991), "Every Minute Counts in Financial Markets," *Journal of International Money and Finance* 10, 23–52.

Gouriéroux, C., A. Monfort and E.Renault (1993), "Indirect Inference," *Journal of Applied Econometrics* 8, Supplement, S85–S118.

Harris., L. (1987), "Transaction Data Tests of the Mixture of Distributions Hypothesis," *Journal of Financial and Quantitative Analysis* 88, 127–141.

Harvey, A.C., E. Ruiz and N. Shephard (1994), "Multivariate Stochastic Variance Models," *Review of Economic Studies* 61, 247–264.

Heston, S.L. (1993), "A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies* 6, 327–343.

Hull, J. and A. White (1987), "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance* 42, 281–300.

Ito, T., and V.V. Roley (1987), "News from the US and Japan: Which Moves the Yen/Dollar Exchange Rate?," *Journal of Monetary Economics* 19, 255–277.

Johnson, H. and D. Shanno (1987), "Option Pricing when the Variance Is Changing," *Journal of Financial and Quantitative Analysis* 22, 143–152.

Karpoff, J. (1987), "The Relation between Price Changes and Trading Volume: A Survey," *Journal of Financial and Quantitative Analysis* 22, 109–126.

Lamoureux, C. and W. Lastrapes (1990), "Heteroskedasticity in Stock Return Data: Volume versus GARCH Effect," *Journal of Finance* XLV(1), 221–229.

Lamoureux, C. and W. Lastrapes (1992), "Forecasting Stock Return Variance: Toward an Understanding of Stock Implied Volatilities," Working Paper, J.M. Ohlin School of Business, Washington University.

Lamoureux, C. and W. Lastrapes (1993), "Endogenous Trading Volume and Momentum in Stock Return Volatility," Working Paper, J.M. Ohlin School of Business, Washington University.

Lo, A.W. (1988), "Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data," *Econometric Theory* 4, 231–247.

Mandelbrot, B. and H. Taylor (1967), "On the Distribution of Stock Prices Differences," *Operations Research* 15, 1057–1062.

Melino, A. (1991), "Estimation of Continuous Time Models in Finance," Working Paper, University of Toronto.

Melino, A. and S.M. Turnbull (1990), "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics* 45(1/2), 239–267.

Miron, J.A. (1994), "The Economics of Seasonal Cycles," in C.A. Sims (ed.) *Advances in Econometrics — Sixth World Congress* (Cambridge University Press).

Moody, J. and L. Wu (1989), "Statistical Analysis and Forecasting of High Frequency Foreign Exchange Rates," Working Paper, Oregon Graduate Institute.

Müller, U.A., M.M. Dacorogna, R.B. Olsen, O.V. Pictet, M. Schwarz and C. Morgenegg (1990), "Statistical Study of Foreign Exchange Rates. Empirical Evidence of a Price Change Scaling Law and Intraday Analysis," *Journal of Banking and Finance* 14, 1189–1208.

Müller, U.A., M.M. Dacorogna, R.D. Davé, O.V. Pictet, R.B. Olsen and J.R. Ward (1993), "Fractals and Intrinsic Time — A Challenge to Econometricians," Discussion Paper O & A, Zürich.

Nelson, D.B. (1988), "Time Series Behavior of Stock Market Volatility and Returns," Ph.D. dissertation, MIT.

Nelson, D.B. (1989), "Modelling Stock Market Volatility Changes," *Proceedings of the American Statistical Association, Business and Economics Section*, 93–98.

Nelson, D.B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica* 59, 347–370.

Pagan, A. and G.W. Schwert (1990), "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, Cambridge University Press, Cambridge.

Poi, J. and W. Polasek (1995), "ISAR–ARCH Model," paper presented at the High Frequency Data Conference, Zürich, Switzerland, March 1995.

Schwert, G.W. (1990), "Indexes of U.S. Stock Prices from 1802 to 1987," *Journal of Business* 63, 399–426.

Scott, L.O. (1987), "Option Pricing when the Variance Changes Randomly: Theory, Estimation and an Application," *Journal of Financial and Quantitative Analysis* 22, 419–438.

Stein, E.M. and J.C. Stein (1991), "Stock Price Distributions with Stochastic Volatility: An Analytic Approach," *Review of Financial Studies* 4, 727–752.

Stock, J.H. (1988), "Estimating Continuous Time Processes Subject of Time Deformation," *Journal of Statistical Association* 83(401), 77–84.

Tauchen, G.E. and M. Pitts (1983), "The Price Variability Volume Relationship on Speculative Markets," *Econometrica* 51(2), 485–505.

Wasserfallen, W. (1989), "Flexible Exchange Rates: A Closer Look," *Journal of Monetary Economics* 23, 511–521.

Wiggins, J. (1987), "Options Values under Stochastic Volatility: Theory and Empirical Evidence," *Journal of Financial Economics* 19, pp. 351–372.

Remerciements

To accomplish this work, the generous support of my thesis supervisors, Jean-Marie Dufour and Eric Ghysels, is acknowledged. I would like to thank both of them for sharing with me their knowledge and ideas. I am very grateful for the time they spent working with me, their commitment, inspiring discussions and guidance. I truly appreciate the opportunity of working under their direction.

I am also grateful to my husband, Haldun Sarlan, for "standing by" in bad and in good moments throughout the entire Ph.D. program, for his collaboration and encouragement. Finally, I would like to say "thank you" to my Mom for everything she's done for me.