Université de Montréal

Problèmes d'économétrie en macroéconomie et en finance: mesures de causalité, asymétrie de la volatilité et risque financier

par

Abderrahim Taamouti

Département de 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

Juin 2007

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée:

Problèmes d'économétrie en macroéconomie et en finance: mesures de causalité, asymétrie de la volatilité et risque financier

présentée par

Abderrahim Taamouti

a été évaluée par un jury composé des personnes suivantes:

William McCausland: président -rapporteur

Jean-Marie Dufour: directeur de recherche

Nour Meddahi: co-directeur de recherche

Marine Carrasco: membre du jury

Emma M. Iglesias: examinateur externe (University of Michigan)

Jean Boivin: représentant du doyen de la FES

Sommaire

Cette thèse de doctorat consiste en quatre essais traitant des problèmes d'économétrie en macroéconomie et en finance. Nous étudions trois sujets principaux : (1) mesures de causalité à différent horizons avec des applications macroéconomiques et financières (essais nos 1 et 2); (2) mesures de risque financier et gestion de portefeuille dans les modèles à changements de régime markoviens (essai no 3); (3) développement de méthodes d'inférence non paramétriques exactes, optimales et adaptatives dans les modèles de régression linéaires et non linéaires, avec des erreurs non gaussiennes et en présence d'hétéroscédasticité de forme inconnue (essai no 4). De brefs résumés de ces quatre essais sont présentés ci-après.

Dans le premier essai, nous proposons des mesures de causalité à des horizons plus grands que un, lesquelles généralisent les mesures de causalité habituelles qui se limitent à l'horizon un. Ceci est motivé par le fait que, en présence d'un vecteur de variables auxiliaires Z, il est possible que la variable Y ne cause pas la variable X à l'horizon un, mais qu'elle cause celle-ci à un horizon plus grand que un [voir Dufour et Renault (1998)]. Dans ce cas, on parle d'une causalité indirecte transmise par la variable auxiliaire Z. Nous proposons une nouvelle approche pour évaluer ces mesures de causalité en simulant un grand échantillon à partir du processus d'intérêt. Des intervalles de confiance non paramétriques, basés sur la technique de bootstrap, sont également proposés. Finalement, nous présentons une application empirique où est analysée la causalité à différents horizons entre la monnaie, le taux d'intérêt, les prix et le produit intérieur bruit aux États-Unis.

Dans le deuxième essai, nous analysons et quantifions la relation entre la volatilité et les rendements en utilisant des données à haute fréquence. Ceci est important pour la gestion de risque ainsi que pour l'évaluation des produits dérivés. Dans le cadre d'un modèle vectoriel linéaire autorégressif de rendements et de la volatilité réalisée, nous quantifions l'effet de levier et l'effet de la volatilité sur les rendements (ou l'effet rétroactif de la volatilité) en se servant des mesures de causalité à court et à long terme proposées dans l'essai 1. En utilisant des observations à chaque 5 minute sur l'indice boursier S&P 500, nous mesurons une faible présence de l'effet de levier dynamique pour les quatre premières heures dans les données horaires et un important effet de levier dynamique pour les trois premiers jours dans les données journalières. L'effet de la volatilité sur les rendements s'avère négligeable et non significatif à tous les horizons. Nous utilisons également ces mesures de causalité pour quantifier et tester l'impact des bonnes et des mauvaises nouvelles sur la volatilité. Empiriquement, nous mesurons un important impact des mauvaises nouvelles, ceci à plusieurs horizons. Statistiquement, l'impact des mauvaises nouvelles est significatif durant les quatre premiers jours, tandis que l'impact de bonnes nouvelles reste négligeable à tous les horizons.

Dans le troisième essai, nous modélisons les rendements des actifs sous forme d'un processus à changements de régime markovien afin de capter les propriétés importantes des marchés financiers, telles que les queues épaisses et la persistance dans la distribution des rendements. De là, nous calculons la fonction de répartition du processus des rendements à plusieurs horizons afin d'approximer la Valeur-à-Risque (VaR) conditionnelle et obtenir une forme explicite de la mesure de risque «déficit prévu» d'un portefeuille linéaire à des horizons multiples. Nous caractérisons la frontière efficiente Moyenne-Variance dynamique des portefeuilles linéaire. En utilisant des observations journalières sur les indices boursiers S&P 500 et TSE 300, d'abord nous constatons que le risque conditionnel (variance ou VaR) des rendements d'un portefeuille optimal, quand est tracé comme fonction de l'horizon, peut augmenter ou diminuer à des horizons intermédiaires et converge vers une constante- le risque inconditionnel- à des horizons suffisamment larges. Deuxièmement, les frontières efficientes à des horizons multiples des portefeuilles optimaux changent dans le temps. Finalement, à court terme et dans 73.56% de l'échantillon, le portefeuille optimal conditionnel a une meilleure performance que le portefeuille optimal inconditionnel.

Dans le quatrième essai, nous dérivons un simple test point optimal basé sur les statistiques de signe dans le cadre des modèles de régression linéaires et non linéaires.

Ce test est exact, robuste contre une forme inconnue d'hétéroscedasticité, ne requiert pas d'hypothèses sur la forme de la distribution et il peut être inversé pour obtenir des régions de confiance pour un vecteur de paramètres inconnus. Nous proposons une approche adaptative basée sur la technique de subdivision d'échantillon pour choisir une alternative telle que la courbe de puissance du test de signe point optimal soit plus proche de la courbe de l'enveloppe de puissance. Les simulations indiquent que quand on utilise à peu près 10% de l'échantillon pour estimer l'alternative et le reste, à savoir 90%, pour calculer la statistique du test, la courbe de puissance de notre test est typiquement proche de la courbe de l'enveloppe de puissance. Nous avons également fait une étude de Monte Carlo pour évaluer la performance du test de signe "quasi" point optimal en comparant sa taille ainsi que sa puissance avec celles de certains tests usuels, qui sont supposés être robustes contre hétéroscedasticité, et les résultats montrent la supériorité de notre test.

Mots clés: séries temporelles; causalité au sens de Granger; causalité indirecte; causalité à des horizons multiples; mesure de causalité; prévisibilité; modèle autorégressive; VAR; bootstrap; Monte Carlo; macroéconomie; monnaie; taux d'intérêt; production; inflation; asymétrie dans la volatilité; l'effet de levier; l'effet rétroactif de la volatilité; données à haut-fréquence; volatilité réalisée; modèle à changement de régime; fonction caractéristique; la distribution de probabilité; valeur-à-risque; déficit prévu; rendement agrégé; la borne supérieure de la valeur-à-risque; portefeuille moyenne-variance; test de signe; test point optimal; modèles linéaires; modèles non linéaires; hétéroscédasticité; inférence exacte; distribution libre; l'enveloppe de puissance; subdivision d'échantillon; approche adaptative; projection.

Summary

This thesis consists of four essays treating the problems of econometrics in macroeconomics and finance. Three main topics are considered: (1) the measurement of causality at different horizons with macroeconomics and financial applications (essays 1 and 2); (2) financial risk measures and asset allocation in the context of Markov switching models (essay 3); (3) exact sign-based optimal adaptive inference in linear and nonlinear regression models in the presence of heteroskedasticity and non-normality of unknown form (essay 4). The four essays are summarized below.

In the first essay, we propose measures of causality at horizons greater than one, as opposed to the more usual causality measures which focus on the horizon one. This is motivated by the fact that, in the presence of a vector Z of auxiliary variables, it is possible that a variable Y does not cause another variable X at horizon 1, but causes it at horizons greater than one [see Dufour and Renault (1998)]. In this case, one has indirect causality transmitted by the auxiliary variable Z. In view of the analytical complexity of the measures, a simple approach based on simulating a large sample from the process of interest is proposed to compute the measures. Valid nonparametric confidence intervals, based on bootstrap techniques, are also derived. Finally, the methods developed are applied to study causality at different horizons between money, federal funds rate, gross domestic product deflator, and gross domestic product, in the U.S.

In the second essay, we analyze and quantify the relationship between volatility and returns for high-frequency equity returns. This is important for asset management as well as for the pricing of derivative assets. Within the framework of a vector autoregressive linear model of returns and realized volatility, leverage and volatility feedback effects are measured by applying the short-run and long-run causality measures proposed in Essay 1. Using 5-minute observations on S&P 500 Index, we measure a weak dynamic leverage effect for the first four hours in hourly data and a strong dynamic leverage effect for the first three days in daily data. The volatility feedback effect is found to be negligible at all horizons. We also use the causality measures to quantify and test statistically the

dynamic impact of good and bad news on volatility. Empirically, we measure a much stronger impact for bad news at several horizons. Statistically, the impact of bad news is found to be significant for the first four days, whereas the impact of good news is negligible at all horizons.

In the third essay, we consider a Markov switching model to capture important features such as heavy tails, persistence, and nonlinear dynamics in the distribution of asset returns. We compute the conditional probability distribution function of multi-horizons returns, which we use to approximate the conditional multi-horizons Value-at-Risk (VaR) and we derive a closed-form solution for the multi-horizons conditional Expected Shortfall. We characterize the dynamic Mean-Variance efficient frontier of the optimal portfolios. Using daily observations on S&P 500 and TSE 300 indices, we first found that the conditional risk (variance and VaR) per period of the multi-horizon optimal portfolio's returns, when plotted as a function of the horizon, may be increasing or decreasing at intermediate horizons, and converges to a constant- the unconditional risk-at long enough horizons. Second, the efficient frontiers of the multi-horizon optimal portfolios are time varying. Finally, at short-term and in 73.56% of the sample the conditional optimal portfolio performs better then the unconditional one.

In the fourth essay, we derive simple sign-based point-optimal test in linear and nonlinear regression models. The test is exact, distribution free, robust against heteroskedasticity of unknown form, and it may be inverted to obtain confidence regions for the vector of unknown parameters. We propose an adaptive approach based on split-sample technique to choose an alternative such that the power curve of point-optimal sign test is close to the power envelope curve. The simulation study shows that when using approximately 10% of sample to estimate the alternative and the rest to calculate the test statistic, the power curve of the point-optimal sign test is typically close to the power envelope curve. We present a Monte Carlo study to assess the performance of the proposed "quasi"-point-optimal sign test by comparing its size and power to those of some common tests which are supposed to be robust against heteroskedasticity. The

results show that our procedure is superior.

Keywords: time series; Granger causality; indirect causality; multiple horizon causality; causality measure; predictability; autoregressive model; VAR; bootstrap; Monte Carlo; macroeconomics; money; interest rates; output; inflation; volatility asymmetry; leverage effect; volatility feedback effect; high-frequency data; realized volatility; Markov switching model; characteristic function; probability distribution; Value-At-Risk; Expected Shortfall; aggregate return; upper bound VaR; Mean-Variance portfolio; sign-based test; point-optimal test; linear models; nonlinear models; heteroskedasticity; exact inference; distribution free; power envelope; sample-split; adaptive approach; projection.

Contents

List of Tables

List of Figures

Remerciements

Je tiens à rendre un vibrant hommage à mon directeur de recherche Jean-Marie Dufour pour sa disponibilité, sa patience, sa contribution à la réalisation de ce travail et surtout à m'encourager à ne pas baisser les bras durant les moments difficiles. Je voudrais également remercier mon co-directeur Nour Meddahi pour son encouragement, ses conseils et pour m'avoir prodiguée des commentaires scientifiques très pertinents.

Je voudrais aussi remercier mon co-auteur René Garcia pour sa collaboration significative dans le deuxième chapitre de cette thèse. Il m'a beaucoup appris et je lui suis reconnaissant.

Je voudrais aussi remercier les différents institutions qui m'ont soutenu par leur financement: centre interuniversitaire de recherche en économie quantitative (CIREQ), conseil de recherche en sciences humaines du Canada (CRSH), MITACS (The mathematics of information technology and complex systems) et centre interuniversitaire de recherche en analyse des organisations (CIRANO).

J'ai également bénéficié des commentaires de plusieurs autres personnes notamment Lynda Khalaf, Lutz Kilian, Bernard-Daniel Solomon et Johan Latulippe. Un grand merci à Benoît Perron pour avoir pris le temps de lire certains chapitres de ma thèse et me faire des commentaires très pertinents.

Enfin un remerciement tout spécial à mes parents et Wassima pour l'encouragement et le soutien moral.

Introduction Générale

Cette thèse de doctorat contient quatre chapitres où nous traitons différents problèmes d'économétrie en macroéconomie et en finance. Étant donné l'importance de la causalité pour la compréhension, la prévision et le contrôle des phénomènes économiques, notre premier objectif est de proposer une approche basée sur le concept de causalité pour quantifier et analyser les relations dynamiques entre variables économiques. Dans le contexte macroéconomique par exemple, cette approche serra très utile pour les décideurs des politiques économiques et monétaires puisqu'elle pourra les aider à prendre leurs décisions tout en se basant sur une meilleure connaissance des effets mutuels qu'exerce chaque variable macroéconomique sur les autres à différents horizons. Un autre exemple est celui de la finance où nous pouvons utiliser l'approche proposée pour identifier la meilleure façon de modéliser la relation entre les rendements des actifs et leur volatilité. Ceci reste crucial pour la gestion de risque ainsi que pour l'évaluation des prix des actifs dérivés. Notre deuxième et dernier objectif est de proposer des mesures pour le risque financier et des tests statistiques qui fonctionnent sous des hypothèses plus réalistes. Nous dérivons des mesures de risque financier qui tiennent compte des effets stylisés qu'on observe sur les marchés financiers tels que les queues épaisses et la persistance dans la distribution des rendements. Nous dérivons également des tests optimaux pour tester les valeurs des paramètres dans les modèles de régression linéaires et non-linéaires. Ces tests restent valides sous des hypothèses distributionnelles faibles.

Dans le premier chapitre, nous développons des mesures de causalité à des horizons plus grands que un, par opposition aux mesures de causalité habituelles qui se concentrent sur l'horizon un. Le concept de causalité introduit par Wiener (1956) et Granger (1969) est actuellement reconnu comme étant la notion de base pour étudier les relations dynamiques entre séries temporelles. Ce concept est défini en terme de prévisibilité à l'horizon un d'une variable X à partir de son propre passé, le passé d'une autre variable Y et possiblement un vecteur Z de variables auxiliaires. En se basant sur Granger (1969), nous définissons la causalité de X vers Y une période à l'avance comme suivant:

Y cause X au sens de Granger si les observations de Y jusqu'à la date $t-1$ peuvent aider à prévoir la valeur de $X(t)$ étant donné les passés de X et de Z jusqu'à la date $t-1$. Plus précisément, on dit que Y cause X au sens de Granger si la variance de l'erreur de prévision de $X(t)$ obtenue en utilisant le passé de Y est plus petite que la variance de l'erreur de prévision de $X(t)$ obtenue sans l'utilisation du passé de Y.

Dufour et Renault (1998) ont généralisé le concept de causalité au sens de Granger (1969) en considérant une causalité à un horizon quelconque (arbitraire) h et une causalité jusqu'à un horizon h, où h est un entier positif qui peut être égale à l'infini ($1 \leq h \leq \infty$). Une telle généralisation est motivée par le fait qu'il est possible, en présence d'un vecteur de variables auxiliaires Z, d'avoir la situation où la variable Y ne cause pas la variable X à l'horizon un, mais qu'elle la cause à un long horizon $h > 1$. Dans ce cas, nous parlons d'une causalité indirect transmet par les variables auxiliaires Z.

L'analyse de Wiener-Granger distingue entre trois types de causalités: deux causalités unidirectionnelles (ou effets rétroactifs) de X vers Y, de Y vers X et une causalité instantanée (ou effet instantané) associée aux corrélations contemporaines. En pratique, il est possible que ces trois types de causalités coexistes, d'où l'importance de trouver un moyen pour mesurer leur dégrées et déterminer la plus importante entre elles. Malheuseument, les tests de causalité qu'on trouve dans la littérature échouent à accomplir cette tache, puisqu'ils nous informent uniquement sur la présence ou l'absence d'une causalité. Geweke (1982, 1984) a étendu le concept de causalité en définissant des mesures pour les effets rétroactifs et l'effet instantanée, qu'on peut décomposer dans le domaine du temps et de la fréquence. Gouriéroux, Monfort et Renault (1987) ont proposé des mesures de causalité basées sur l'information de Kullback. Polasek (1994) montre comment des mesures de causalité peuvent être calculées à partir du critère d'information d'Akaike (AIC). Polasek (2000) a aussi introduit des nouvelles mesures de causalité dans le contexte des modèles ARCH univariés et multivariés et leurs extensions en se basant sur l'approche Bayesian.

Les mesures de causalité existantes sont établit seulement pour l'horizon un et échouent donc à capter les effets indirects. Dans le premier chapitre, nous développons des mesures de causalité à différents horizons capables de capter les effets indirects qui apparaissent à des horizons longs. Plus spécifiquement, nous proposons des généralisations à n'importe qu'il horizon h des mesures à l'horizon un proposées par Geweke (1982). Par analogie à Geweke (1982, 1984), nous définissons une mesure de dépendance à l'horizon h qui se décompose en somme des mesures des effets rétroactifs de X vers Y, de Y vers X et de l'effet instantané à l'horizon h.

Pour calculer les mesures associées à un modèle donnée – quand les formules analytiques sont difficiles à obtenir – nous proposons une nouvelle approche basée sur une longue simulation du processus d'intérêt. Pour l'implémentation empirique, nous proposons des estimateurs consistent ainsi que des intervalles de confiance non paramétriques, basés sur la technique de bootstrap. Les mesures de causalité proposées sont appliquées pour étudier la causalité à différents horizons entre la monnaie, le taux d'intérêt, le niveau des prix et la production, aux États-Unis.

Dans le deuxième chapitre, nous mesurons et analysons la relation dynamique entre la volatilité et les rendements en utilisant des données à haute fréquence. Un des effets stylisés caractérisant les marchés financiers est nommé l'asymétrie de la volatilité et signifie que la volatilité à tendance à augmenter plus quand il y a des rendements négatifs que quand il y a des rendements positifs. Dans la littérature il y a deux explications pour ce phénomène. La première est liée à ce qu'on appelle l'effet de levier. Un décroissement dans le prix d'un actif accroit le levier financier et la probabilité de faillite, ce qui rend l'actif risqué et augmentera sa volatilité future [voir Black (1976) et Christie (1982)]. La deuxième explication, ou l'effet rétroactif de la volatilité, est lié à la théorie sur la prime de risque: si la volatilité est évaluée, un accroissement anticipé de celle-ci doit accroître le taux de rendement, ce qui exige un déclin immédiat du prix de l'actif pour permettre un accroissement du rendement futur [voir Pindyck (1984), French, Schwert et Stambaugh (1987), Campbell et Hentschel (1992) et Bekaert et Wu (2000)].

Bekaert et Wu (2000) et récemment Bollerslev et al. (2005), ont fait remarquer que la différence entre les deux explications de l'asymétrie de la volatilité est liée à la question de causalité. L'effet de levier explique pourquoi un rendement négatif conduira à une augmentation future de la volatilité, tandis que l'effet rétroactif de la volatilité justifie comment une augmentation de la volatilité peut conduire à un rendement négatif. Ainsi, l'asymétrie de la volatilité peut être le résultat de divers liens causals: des rendements vers la volatilité, de la volatilité vers les rendements, d'une causalité instantanée, de tous ces effets causals ou de certains d'entre eux.

Bollerslev et al. (2005) ont étudié ces relations en utilisant des données à haut fréquence et des mesures de la volatilité réalisée. Cette stratégie augmente les chances de détecter les vrais liens causals puisque l'agrégation peut rendre la relation entre les rendements et la volatilité simultanée. Leur approche empirique consiste alors à utiliser des corrélations entre les rendements et la volatilité réalisée pour mesurer et comparer la magnitude de l'effet de levier ou de l'effet rétroactif de la volatilité. Cependant, la corrélation est une mesure d'une association linéaire et n'implique pas nécessairement une relation causale. Dans le deuxième chapitre, nous proposons une approche qui consiste à utiliser des données à haut fréquence, modéliser les rendements et la volatilité sous forme d'un modèle vectoriel autorégressif (VAR) et utiliser les mesures de causalité à court et à long terme proposées dans le premier chapitre, pour quantifier et comparer l'effet rétroactif de la volatilité et l'effet de levier dynamiques.

Les études concentrées sur l'hypothèse d'un effet de levier [voir Christie (1982) et Schwert (1989)] ont conclu qu'on ne peut pas compter uniquement sur le changement de la volatilité. Cependant, pour l'effet rétroactif de la volatilité on trouve des résultats empiriques contradictoires. French, Schwert et Stambaugh (1987) et Campbell et Hentschel (1992) ont conclu que la relation entre la volatilité et les rendements espérés à l'horizon un est positive, tandis que Turner, Startz et Nelson (1989), Glosten, Jagannathen et Runkle (1993) et Nelson (1991) ont trouvé que cette relation est négative. Plus souvent le coefficient reliant la volatilité aux rendements est statistiquement non significatif. Pour des actifs individuels, Bekaert et Wu (2000) montrent empiriquement que l'effet rétroactif de la volatilité domine l'effet de levier. En utilisant des données à haut fréquence, Bollerslev et al. (2005) ont obtenu une corrélation négative entre la volatilité et les rendements courants et retardés pour plusieurs jours. Cependant, les corrélations entre les rendements et les retards de la volatilité sont tous proches de zéro.

La deuxième contribution du chapitre deux consiste à montrer que les mesures de causalité peuvent être utilisées pour quantifier l'impact dynamique des bonnes et des mauvaises nouvelles sur la volatilité. L'approche commune pour visualiser empiriquement la relation entre les nouvelles et la volatilité est fournie par la courbe de l'impact des nouvelles originalement étudiée par Pagan et Schwert (1990) et Engle et Ng (1993). Pour étudier l'effet des chocs courants des rendements sur la volatilité espérée, Engle et Ng (1993) ont introduit ce qu'ils appellent la fonction d'impact des nouvelles (ci-après FIN). L'idée de base consiste à conditionner à la date $t+1$ par rapport à l'information disponible à la date t, et de considérer alors l'effet d'un choc de rendement à la date t sur la volatilité à la date $t+1$ en isolation. Dans ce chapitre, nous proposons une nouvelle courbe de l'impact des nouvelles sur la volatilité basée sur les mesures de causalité. Contrairement à FIN d'Engle et Ng (1993), notre courbe peut être construite pour des modèles paramétriques et stochastiques de la volatilité, elle tient compte de toute l'information passée de la volatilité et des rendements, et elle concerne des horizons multiples. En outre, nous construisons des intervalles de confiance autour de cette courbe en se basant sur la technique de bootstrap, ce qui fourni une amélioration, en terme d'inférence statistique, par rapport aux procédures courantes.

Dans le troisième chapitre, nous nous intéressons aux mesures de risque financier et à la gestion de portefeuille dans le contexte des modèles à changements de régime markoviens. Depuis le travail séminal de Hamilton (1989), les modèles à changement de régime sont utilisés de plus en plus en économétrie financière et séries temporelles. Cela à cause de leur capacité à capter certaines importantes caractéristiques, telle que les queues épaisses, la persistance et la dynamique non linéaire dans la distribution des rendements des actifs. Dans ce chapitre, nous exploitons la supériorité de ces modèles pour dériver des mesures de risque financier, tels que la Valeur-à-Risque (VaR) et le déficit prévu, qui tiennent compte des effets stylisés observés sur les marchés financiers. Nous caractérisons également la frontière efficiente moyenne-variance de portefeuilles linéaires à des horizons multiples et nous comparons la performance d'un portefeuille optimal conditionnel et celle d'un portefeuille optimal inconditionnel.

La VaR est devenu la technique la plus utilisée pour mesurer et contrôler le risque sur les marchés financiers. C'est une mesure quantile qui quantifie la perte maximale espérée sur un certain horizon (typiquement une journée ou une semaine) et à un certain niveau de confiance (typiquement 1%, 5% ou 10%). Il existe différentes méthodes pour estimer la VaR sous différents modèles des facteurs de risque. Généralement, il y a un arbitrage à faire entre la simplicité de la méthode d'estimation et le réalisme des hypothèses dans le modèle des facteurs de risque considéré: Plus on permet à ce dernier de capter plus d'effets stylisés, plus la méthode d'estimation devient complexe. Sous l'hypothèse de la normalité des rendements, nous pouvons montrer que la VaR est donné par une simple formule analytique [voir RiskMetrics (1995)]. Cependant, quand nous relâchons cette hypothèse, le calcul analytique de la VaR devient plus compliqué et les gens ont tendance à utiliser des méthodes de simulations. En se basant sur un modèle à changement de régime, ce chapitre propose des approximations analytiques de la VaR conditionnelle sous des hypothèses plus réalistes comme la non normalité des rendements.

L'estimation de la VaR dans le contexte des modèles à changement de régime à été abordée par Billio et Pelizzon (2000) et Guidolin et Timmermann (2004). Billio et Pelizzon (2000) ont utilisé un modèle de volatilité avec changement de régime pour prévoir la distribution des rendements et estimer la VaR des actifs simples et des portefeuilles linéaires. En comparant la VaR calculée à partir d'un modèle à changement de régime avec celles obtenues en utilisant l'approche variance-covariance ou un GARCH(1,1), ils ont conclu que la VaR d'un modèle à changement de régime est préférable à celles des autres approches. Guidolin et Timmermann (2004) ont examiné la structure à terme de la VaR sous différents modèles économétriques, y compris les modèles à changement de régime multivariés, et ils ont constaté que le bootstrap et les modèles à changement de régime sont les meilleurs, parmi d'autres modèles considérés, pour estimer des VaR à des niveaux de 5% et 1%, respectivement. À notre connaissance, aucune méthode analytique n'a été proposée pour estimer la VaR conditionnelle dans le contexte des modèles à changement de régime. Ce chapitre utilise la même approche que Cardenas et autres (1997), Rouvinez (1997) et Duffie et Pan (2001) pour fournir une approximation analytique de la VaR conditionnelle à des horizons multiples. D'abord, en utilisant la méthode d'inversion de Fourier, nous dérivons la fonction de répartition des rendements de portefeuilles linéaires à des horizons multiples. Ensuite, pour rendre l'estimation de la VaR possible nous employons une méthode numérique d'intégration, conçue par Davies (1980), pour approximer l'opérateur intégrale dans la formule d'inversion. Finalement, nous utilisons le filtre d'Hamilton pour estimer la VaR conditionnelle.

En dépit de sa popularité parmi les gestionnaires et les régulateurs, la VaR a été critiquée du fait, qu'en général, elle n'est pas consistante et ignore des pertes au delà de son niveau. En outre, elle n'est pas subadditive, ce qui signifie qu'elle pénalise la diversification au lieu de la récompenser. En conséquence, des chercheurs ont proposé une nouvelle mesure de risque, appelée le déficit prévu, qui est égale à l'espérance conditionnelle de la perte étant donné que celle-ci est au delà du niveau de VaR. Contrairement à la VaR, la mesure déficit prévu est consistante, tient compte de la fréquence et la sévérité des pertes financières, et elle est additive. À notre connaissance, aucune formule analytique n'a été dérivée pour la mesure déficit prévue conditionnelle dans le contexte des modèles à changement de régime. Dans ce chapitre nous proposons une solution explicite de cette mesure pour des portefeuilles linéaires et à des horizons multiples.

Un autre objectif du chapitre trois est d'étudier le problème de gestion de portefeuille dans le contexte des modèles à changement de régime. Dans la littérature il y a deux manières de considérer le problème d'optimisation de portefeuille: statique et dynamique. Dans le contexte moyenne-variance, la différence entre les deux est liée à comment nous calculons les deux premiers moments des rendements. Dans l'approche statique, la structure de portefeuille optimale est choisie une fois pour toute au début de la période. Un inconvénient de cette approche c'est qu'elle suppose une moyenne et une variance des rendements constantes. Dans l'approche dynamique, la structure du portefeuille optimale est continuellement ajustée en utilisant l'ensemble d'information observé à la date courante. Un avantage de cette approche est qu'elle permet d'exploiter la prévisibilité des deux premiers moments pour bien gérer les opportunités d'investissement.

Des études récentes qu'ont examiné les implications économiques de la prévisibilité des rendements sur la gestion de portefeuille ont constaté que les investisseurs agissent différemment quand les rendements sont prévisibles. Nous distinguons entre deux approches. La première, qui évalue les avantages économiques par l'intermédiaire du calibrage antérieur, conclut que la prévisibilité des rendements améliore les décisions des investisseurs [voir Kandel et Stambaugh (1996), Balduzzi et Lynch (1999), Lynch (2001), Gomes (2002) et Campbell, Chan et Viceira (2002)]. La deuxième approche, qui évalue la performance postériori de la prévisibilité des rendements, trouve des résultats différents. Breen, Glosten et Jagannathan (1989) et Pesaran et Timmermann (1995), ont constaté que la prévisibilité rapporte des gains économiques significatifs hors échantillon, tandis que Cooper, Gutierrez et Marcum (2001) et Cooper et Gulen (2001) n'ont constaté aucun gain économique significatif. Dans le contexte moyenne-variance, Jacobsen (1999) et Marquering et Verbeek (2001) ont constaté que les gains économiques de l'exploitation de la prévisibilité des rendements sont significatifs, alors que Handa et Tiwari (2004) ont conclut que ces gains sont incertains.[1]

Récemment, Campbell et Viceira (2005) ont examiné les implications à des horizons multiples de la prévisibilité pour un portefeuille moyenne-variance en utilisant un modèle vectoriel autorégressif standard avec une matrice variance-covariance constante pour les termes d'erreurs. Ils ont conclut que les changements dans les opportunités d'investissements peuvent alterner l'arbitrage rendement-risque pour les obligations, les actifs et l'argent comptant à travers les horizons d'investissement et que la prévisibilité a

[1] Voir Han (2005) pour plus de discussion.

des effets importants sur la structure de la variance et les corrélations des actifs à travers les horizons d'investissement. Dans le troisième chapitre nous étendons le modèle de Campbell et Viceira (2005) en considérant un modèle à changement de régimes. Cependant, nous ne tenons pas compte des variables, telles que le prix-revenu, le taux d'intérêt et d'autres, pour prévoir les rendements futurs, comme dans Campbell et Viceira (2005). Nous dérivons les deux premiers moments conditionnels et inconditionnels à des horizons multiples, que nous utilisons pour comparer la performance des portefeuilles optimaux conditionnels et inconditionnels. En utilisant des observations journalières sur les indices boursiers S&P 500 et TSE 300, d'abord nous constatons que le risque conditionnel (variance ou VaR) des rendements d'un portefeuille optimal, quand est tracé comme fonction de l'horizon h, peut augmenter ou diminuer à des horizons intermédiaires et converge vers une constante- le risque inconditionnel- à des horizons suffisamment larges. Deuxièmement, les frontières efficientes à des horizons multiples des portefeuilles optimaux changent dans le temps. Finalement, à court terme et dans 73.56% de l'échantillon, le portefeuille optimal conditionnel a une meilleure performance que le portefeuille optimal inconditionnel.

Dans le quatrième et dernier chapitre, nous développons des méthodes d'inférences exactes et non paramétriques dans le contexte des modèles de régression linéaires et non linéaires. En pratique, la plupart des données économiques sont hétéroscédastique et non normale. En présence de quelques formes d'hétéroscédasticité, les tests paramétriques proposés pour améliorer l'inférence peuvent ne pas contrôler le niveau et avoir une puissance faible. Par exemple, quand il y a des sauts dans la variance des termes erreurs, nos résultats de simulation indiquent que les tests statistiques habituels basés sur la correction de la variance proposée par White (1980), qui sont censés être robuste contre l'hétéroscédasticité, ont une puissance très faible. D'autres formes d'hétéroscédasticité pour lesquelles les tests habituels sont moins puissants sont une variance exponentielle et un GARCH avec un ou plusieurs valeurs aberrantes. En même temps, beaucoup de tests paramétriques exacts développés dans la littérature supposent typiquement que les termes d'erreurs sont normaux. Cette hypothèse est peu réaliste et en présence de distributions avec des queues épaisses et/ou asymétriques, nos résultats de simulation montrent que ces tests peuvent ne pas contrôler le niveau et avoir de la puissance. En outre, les procédures statistiques développées pour faire de l'inférence sur des paramètres de modèles non-linéaires sont typiquement basées sur des approximations asymptotiques. Cependant, ces derniers peuvent être invalides même dans de grands échantillons [voir Dufour(1997)]. Ce chapitre à pour objectif de proposer des procédures statistiques exactes qui fonctionnent sous des hypothèses plus réalistes. Nous dérivons des tests optimaux basés sur les statistiques de signe pour tester les valeurs des paramètres dans les modèles de régression linéaires et non-linéaires. Ces tests sont valides sous des hypothèses distributionnelles faibles telles que l'hétéroscédasticité de forme inconnue et la non-normalité.

Plusieurs auteurs ont fourni des arguments théoriques pour justifier pourquoi les tests paramétriques existants pour tester la moyenne des observations i.i.d. échouent sous des hypothèses distributionnelles faibles telles que la non-normalité et l'hétéroscédasticité de forme inconnue. Bahadur et Savage (1956) ont montré que sous des hypothèses distributionnelles faibles sur les termes d'erreurs, il est impossible d'obtenir un test valide pour la moyenne d'observations i.i.d. même pour de grands échantillons. Plusieurs autres hypothèses au sujet de divers moments des observations i.i.d. conduisent à des difficultés semblables et ceci peut être expliqué par le fait que les moments ne sont pas empiriquement significatifs dans les modèles non paramétriques ou des modèles avec des hypothèses faibles. Lehmann et Stein (1949) et Pratt et Gibbons (1981, sec. 5.10) ont prouvé que les méthodes de signe conditionnelles étaient la seule manière possible de produire des procédures d'inférence exactes dans des conditions d'hétéroscédasticité de forme inconnue et de la non-normalité. Pour plus de discussion au sujet des problèmes d'inférence statistiques dans les modèles non paramétriques le lecteur peut consulter Dufour (2003).

Dans ce chapitre nous introduisons de nouveaux tests basés sur les statistiques de signe pour tester les valeurs des paramètres dans des modèles de régression linéaires et non-linéaire. Ces tests sont exacts, n'exigent pas de spécifier la distribution des termes d'erreurs, robuste contre une hétéroscédasticité de forme inconnue, et ils peuvent être inversés pour obtenir des régions de confiance pour un vecteur de paramètres inconnus. Ces tests sont dérivés sous les hypothèses que les termes d'erreurs dans le modèle de régression sont indépendants, et non nécessairement identiquement distribués, avec zéro médian conditionnellement aux variables explicatives. Seulement quelques procédures de test de signe ont été développées dans la littérature. En présence d'une seule variable explicative, Campbell et Dufour (1995, 1997) ont proposé des analogues non paramétriques du t-test, basés sur des statistiques de signe et de rang, qui sont applicables à une classe spécifique de modèles rétroactifs comprenant le modèle de Mankiw et Shapiro (1987) et le modèle de marche aléatoire. Ces tests sont exacts même si les perturbations sont asymétriques, non-normales, et hétéroscédastique. Boldin, Simonova et Tyurin (1997) ont proposé des procédures d'inférence et d'estimation localement optimales dans le contexte des modèles linéaires basés sur les statistiques de signes. Coudin et Dufour (2005) ont étendu le travail de Boldin et al (1997) à la présence de certain formes de dépendance statistique dans les données. Wright (2000) a proposé des tests de ratio de variance basés sur les rangs et les signes pour tester l'hypothèse nulle que la série d'intérêt est une séquence de différence martingale.

Dans ce chapitre nous abordons la question d'optimalité et nous cherchons à dériver des tests point-optimaux basés sur les statistiques de signes. Les tests point-optimaux sont utiles dans plusieurs directions et ils sont plus attractives pour les problèmes dans lesquels l'espace de paramètre peut être limité par des considérations théoriques. En raison de leurs propriétés de puissance, les tests point-optimaux sont particulièrement attractive quand on teste une théorie économique contre une autre, par exemple une nouvelle théorie contre un autre qui existe déjà. Ils ont une puissance optimale à un point donné et, dépendamment de la structure du problème, pourrait avoir une puissance optimale pour l'ensemble de l'espace de paramètres. Une autre caractéristique intéressante des tests point-optimaux c'est qu'ils peuvent être utilisés pour tracer l'enveloppe de puissance pour un problème de test donné. Cette enveloppe de puissance fournit un repère évident contre lequel des procédures de test peuvent être évaluées. Pour plus de discussion concernant l'utilité des tests point-optimaux le lecteur peut consulter King (1988). Plusieurs auteurs ont dérivé des tests point-optimaux pour améliorer l'inférence pour quelques problèmes économiques. Dufour et King (1991) ont utilisés les tests point-optimaux pour tester le coefficient d'autocorrélation d'un modèle de régression linéaire avec des termes d'erreurs normales autorégressives d'ordre un. Elliott, Rothenberg, et Stock (1996) ont dérivé l'enveloppe de puissance asymptotique pour des tests point-optimaux d'une racine unitaire dans la représentation autorégressive d'une série temporelle gaussienne sous différentes formes de tendance. Plus récemment, Jansson (2005) a dérivé une enveloppe de puissance gaussienne asymptotique pour des tests de l'hypothèse nulle du cointegration et a proposé un test point-optimal faisable de cointegration dont la fonction de puissance asymptotique locale s'avère proche de l'enveloppe de puissance gaussienne asymptotique.

Puisque notre test point optimal dépend de l'hypothèse alternative, nous proposons une approche adaptative basée sur la technique de la subdivision de l'échantillon [voir Dufour et Torres(1998) et Dufour et Jasiak (2001)] pour choisir une alternative tels que la courbe de puissance du test de signe point-optimal est proche de celle de l'enveloppe de puissance. L'étude de simulation montre qu'en utilisant approximativement 10% de l'échantillon pour estimer l'alternative et le reste, voir 90%, pour calculer la statistique de test, la courbe de puissance du test de signe point-optimal est typiquement proche de la courbe d'enveloppe de puissance. En faisant une étude de Monte Carlo pour évaluer la performance du test de signe quasi-point-optimal en comparant sa taille et sa puissance à celles de quelques tests communs qui sont censés être robustes contre l'hétéroscédascticité. Les résultats prouvent que les procédures adaptatives de signe semblent être supérieures.

Chapter 1

Short and long run causality measures: theory and inference

1.1 Introduction

The concept of causality introduced by Wiener (1956) and Granger (1969) is now a basic notion for studying dynamic relationships between time series. This concept is defined in terms of predictability at horizon one of a variable X from its own past, the past of another variable Y, and possibly a vector Z of auxiliary variables. Following Granger (1969), we define causality from Y to X one period ahead as follows: Y causes X if observations on Y up to time $t-1$ can help to predict $X(t)$ given the past of X and Z up to time $t-1$. More precisely, we say that Y causes X if the variance of the forecast error of X obtained by using the past of Y is smaller than the variance of the forecast error of X obtained without using the past of Y.

The theory of causality has generated a considerable literature. In the context of bivariate ARMA models, Kang (1981), derived necessary and sufficient conditions for noncausality. Boudjellaba, Dufour, and Roy (1992, 1994) developed necessary and sufficient conditions of noncausality for multivariate ARMA models. Parallel to the literature on noncausality conditions, some authors developed tests for the presence of causality between time series. The first test is due to Sims (1972) in the context of bivariate time series. Other tests were developed for VAR models [see Pierce and Haugh (1977), Newbold (1982), Geweke (1984a)] and VARMA models [see Boudjellaba, Dufour, and Roy (1992, 1994)].

In Dufour and Renault (1998) the concept of causality in the sense of Granger (1969) is generalized by considering causality at a given (arbitrary) horizon h and causality up to horizon h, where h is a positive integer and can be infinite $(1 \leq h \leq \infty)$; for related work, see also Sims (1980), Hsiao (1982), and Lütkepohl (1993). Such generalization is motivated by the fact that, in the presence of auxiliary variables Z, it is possible to have the variable Y not causing variable X at horizon one, but causing it at a longer horizon

$h > 1$. In this case, we have an indirect causality transmitted by the auxiliary variables Z. Necessary and sufficient conditions of noncausality between vectors of variables at any horizon h for stationary and nonstationary processes are also supplied.

The analysis of Wiener-Granger distinguishes among three types of causality: two unidirectional causalities (called feedbacks) from X to Y and from Y to X and an instantaneous causality associated with contemporaneous correlations. In practice, it is possible that these three types of causality coexist, hence the importance of finding means to measure their degree and determine the most important ones. Unfortunately, existing causality tests fail to accomplish this task, because they only inform us about the presence or the absence of causality. Geweke (1982, 1984) extended the causality concept by defining measures of feedback and instantaneous effects, which can be decomposed in time and frequency domains. Gouriéroux, Monfort, and Renault (1987) proposed causality measures based on the Kullback information. Polasek (1994) showed how causality measures can be calculated using the Akaike Information Criterion (AIC). Polasek (2000) also introduced new causality measures in the context of univariate and multivariate ARCH models and their extensions based on a Bayesian approach.

Existing causality measures have been established only for a one period horizon and fail to capture indirect causal effects. In this chapter, we develop measures of causality at different horizons which can detect the well known indirect causality that appears at higher horizons. Specifically, we propose generalizations to any horizon h of the measures proposed by Geweke (1982) for the horizon one. Both nonparametric and parametric measures for feedback and instantaneous effects at any horizon h are studied. Parametric measures are defined in terms of impulse response coefficients of the moving average (MA) representation of the process. By analogy with Geweke (1982, 1984), we also define a measure of dependence at horizon h which can be decomposed into the sum of feedback measures from X to Y, from Y to X, and an instantaneous effect at horizon h. To evaluate the measures associated with a given model – when analytical formulae are difficult to obtain – we propose a new approach based on a long simulation of the process of interest.

For empirical implementation, we propose consistent estimators as well as nonparametric confidence intervals, based on the bootstrap technique. The proposed causality measures can be applied in different contexts and may help to solve some puzzles of the economic and financial literature. They may improve the well known debate on long-term predictability of stock returns. In the present chapter, they are applied to study causality relations at different horizons between macroeconomic, monetary and financial variables in the U.S. The data set considered is the one used by Bernanke and Mihov (1998) and Dufour, Pelletier, and Renault (2006). This data set consists of monthly observations on nonborrowed reserves, the federal funds rate, gross domestic product deflator, and real gross domestic product.

The plan of this chapter is as follows. Section 1.2 provides the motivation behind an extension of causality measures at horizon $h > 1$. Section 1.3 presents the framework allowing the definition of causality at different horizons. In section 1.4, we propose nonparametric short-run and long-run causality measures. In section 1.5, we give a parametric equivalent, in the context of linear stationary invertible processes, of the causality measures suggested in section 1.4. Thereafter, we characterize our measures in the context of moving average models with finite order q. In section 1.6 we discuss different estimation approaches. In section 1.7 we suggest a new approach to calculate these measures based on the simulation. In section 1.8 we establish the asymptotic distribution of measures and the asymptotic validity of their nonparametric bootstrap confidence intervals. Section 1.9 is devoted to an empirical application and the conclusion relating to the results is given in section 1.10. Technical proof are given in section 1.11.

1.2 Motivation

The causality measures proposed in this chapter constitute extensions of those developed by Geweke (1982, 1984) and others [see the introduction]. The existing causality measures quantify the effect of a vector of variables on another at one period horizon. The significance of such measures is however limited in the presence of auxiliary variables, since it is possible that a vector Y causes another vector X at horizon h strictly higher than 1 even if there is no causality at horizon 1. In this case, we speak about an indirect effect induced by the auxiliary variables Z. Clearly causality measures defined for horizon 1 are unable to quantify this indirect effect. This chapter proposes causality measures at different horizons to quantify the degree of short and long run causality between vectors of random variables. Such causality measures detect and quantify the indirect effects due to auxiliary variables. To illustrate the importance of such causality measures, consider the following examples.

Example 1 *Suppose we have information about two variables X and Y. $(X, Y)'$ is a stationary VAR(1) model:*

$$\begin{bmatrix} X(t+1) \\ Y(t+1) \end{bmatrix} = \begin{bmatrix} 0.5 & 0.7 \\ 0.4 & 0.35 \end{bmatrix} \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} + \begin{bmatrix} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \end{bmatrix}. \tag{1.1}$$

$X(t+1)$ is given by the following equation:

$$X(t+1) = 0.5\ X(t) + 0.7\ Y(t) + \varepsilon_X(t+1). \tag{1.2}$$

Since the coefficient of $Y(t)$ in (1.2) is equal to 0.7, we can conclude that Y causes X in the sense of Granger. However, this does not give any information on causality at horizons larger than 1 nor on its strength. To study the causality at horizon 2, let us consider the system (1.1) at time $t+2$:

$$\begin{bmatrix} X(t+2) \\ Y(t+2) \end{bmatrix} = \begin{bmatrix} 0.53 & 0.595 \\ 0.34 & 0.402 \end{bmatrix} \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} + \begin{bmatrix} 0.5 & 0.7 \\ 0.4 & 0.35 \end{bmatrix} \begin{bmatrix} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \end{bmatrix} + \begin{bmatrix} \varepsilon_X(t+2) \\ \varepsilon_Y(t+2) \end{bmatrix}.$$

In particular, $X(t+2)$ is given by

$$X(t+2) = 0.53\ X(t) + 0.595Y(t) + 0.5\varepsilon_X(t+1) + 0.7\varepsilon_Y(t+1) + \varepsilon_X(t+2)\,. \tag{1.3}$$

The coefficient of $Y(t)$ in equation (1.3) is equal to 0.595, so we can conclude that Y causes X at horizon 2. But, how can one measure the importance of this "long-run" causality? Existing measures do not answer this question.

Example 2 *Suppose now that the information set contains not only the two variables of interest X and Y but also an auxiliary variable Z. We consider a trivariate stationary process $(X, Y, Z)'$ which follows a VAR(1) model:*

$$\begin{bmatrix} X(t+1) \\ Y(t+1) \\ Z(t+1) \end{bmatrix} = \begin{bmatrix} 0.60 & 0 & 0.80 \\ 0 & 0.40 & 0 \\ 0 & 0.60 & 0.10 \end{bmatrix} \begin{bmatrix} X(t) \\ Y(t) \\ Z(t) \end{bmatrix} + \begin{bmatrix} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \\ \varepsilon_Z(t+1) \end{bmatrix}, \tag{1.4}$$

hence

$$X(t+1) = 0.6\ X(t) + 0.8\ Z(t) + \varepsilon_X(t+1)\,. \tag{1.5}$$

Since the coefficient of $Y(t)$ in equation (1.5) is 0, we can conclude that Y does not cause X at horizon 1. If we consider model (1.4) at time $t+2$, we get:

$$\begin{bmatrix} X(t+2) \\ Y(t+2) \\ Z(t+2) \end{bmatrix} = \begin{bmatrix} 0.60 & 0.00 & 0.80 \\ 0.00 & 0.40 & 0.00 \\ 0.00 & 0.60 & 0.10 \end{bmatrix}^2 \begin{bmatrix} X(t) \\ Y(t) \\ Z(t) \end{bmatrix} \tag{1.6}$$

$$+ \begin{bmatrix} 0.60 & 0.00 & 0.80 \\ 0.00 & 0.40 & 0.00 \\ 0.00 & 0.60 & 0.10 \end{bmatrix} \begin{bmatrix} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \\ \varepsilon_Z(t+1) \end{bmatrix} + \begin{bmatrix} \varepsilon_X(t+2) \\ \varepsilon_Y(t+2) \\ \varepsilon_Z(t+2) \end{bmatrix}. \tag{1.7}$$

so that $X(t+2)$ is given by

$$X(t+2) = 0.36\ X(t) + 0.48Y(t) + 0.56\ Z(t) + 0.6\varepsilon_X(t+1) + 0.8\varepsilon_Z(t+1) + \varepsilon_X(t+2). \quad (1.8)$$

The coefficient of $Y(t)$ in equation (1.8) is equal to 0.48, which implies that Y causes X at horizon 2. This shows that the absence of causality at $h = 1$ does not exclude the possibility of a causality at horizon $h > 1$. This indirect effect is transmitted by the variable Z:

$$Y \underbrace{\rightarrow}_{0.6} Z \underbrace{\rightarrow}_{0.8} X,$$

$0.48 = 0.60 \times 0.80$, *where 0.60 and 0.80 are the coefficients of the one period effect of Y on Z and the one period effect of Z on X, respectively. So, how can one measure the importance of this indirect effect? Again, existing measures do not answer this question.*

1.3 Framework

The notion of noncausality considered here is defined in terms of orthogonality conditions between subspaces of a Hilbert space of random variables with finite second moments. We denote $L^2 \equiv L^2(\Omega, \mathcal{A}, Q)$ the Hilbert space of real random variables defined on a common probability space $(\Omega, \mathcal{A}, Q)$, with covariance as inner product.

We consider three multivariate stochastic processes $\{X(t) : t \in \mathbb{Z}\}$, $\{Y(t) : t \in \mathbb{Z}\}$, and $\{Z(t) : t \in \mathbb{Z}\}$, with

$$\begin{aligned} X(t) &= (x_1(t), \ldots, x_{m_1}(t))', \quad x_i(t) \in L^2, \quad i = 1, \ldots, m_1, \\ Y(t) &= (y_1(t), \ldots, y_{m_2}(t))', \quad y_i(t) \in L^2, \quad i = 1, \ldots, m_2, \\ Z(t) &= (z_1(t), \ldots, z_{m_3}(t))', \quad z_i(t) \in L^2, \quad i = 1, \ldots, m_3, \end{aligned}$$

where $m_1 \geq 1$, $m_2 \geq 1$, $m_3 \geq 0$; and $m_1+m_2+m_3 = m$. We denote $\underline{X}_t = \{X(s) : s \leq t\}$,

$\underline{Y}_t = \{Y(s) : s \leq t\}$ and $\underline{Z}_t = \{Z(s) : s \leq t\}$ the information sets which contain all the past and present values of X, Y and Z, respectively. We denote I_t the information set which contains $\underline{X}_t$, $\underline{Y}_t$ and $\underline{Z}_t$. $I_t - A_t$, with $A_t = \underline{X}_t, \underline{Y}_t$ or $\underline{Z}_t$, contains all the elements of I_t except those of A_t. These information sets can be used to predict the value of X at horizon h, denoted $X(t+h)$, for all $h \geq 1$.

For any information set B_t, let $P[x_i(t+h) \mid B_t]$ be the best linear forecast of $x_i(t+h)$ based on the information set B_t. the corresponding prediction error is

$$u\big(x_i(t+h) \mid B_t\big) = x_i(t+h) - P[x_i(t+h) \mid B_t]$$

and $\sigma^2\big(x_i(t+h) \mid B_t\big)$ is the variance of this prediction error. Thus, the best linear forecast of $X(t+h)$ is

$$P(X(t+h) \mid B_t) = \big(P(x_1(t+h) \mid B_t), \ldots, P(x_{m_1}(t+h) \mid B_t)\big)',$$

the corresponding vector of prediction error is

$$U(X(t+h) \mid B_t) = \big(u(x_1(t+h) \mid B_t)', \ldots, u(x_{m_1}(t+h) \mid B_t)\big)'.$$

and its variance-covariance matrix is $\Sigma\big(X(t+h) \mid B_t\big)$. Each component $P[x_i(t+h) \mid B_t]$ of $P[X(t+h) \mid B_t]$, for $1 \leq i \leq m_1$, is then the orthogonal projection of $x_i(t+h)$ on the subspace B_t.

Following Dufour and Renault (1998), noncausality at horizon h and up to horizon h, where h is a positive integer, are defined as follows.

Definition 1 *For $h \geq 1$,*

(i) Y does not cause X at horizon h given $I_t - \underline{Y}_t$, denoted $Y \underset{h}{\nrightarrow} X \mid I_t - \underline{Y}_t$ iff

$$P[X(t+h) \mid I_t - \underline{Y}_t] = P[X(t+h) \mid I_t], \ \forall t > w,$$

where w represents a "starting point":[1]

(ii) Y does not cause X up to horizon h given $I_t - Y_t$, denoted $Y \underset{(h)}{\nrightarrow} X \mid I_t - Y_t$ *iff*

$$Y \underset{k}{\nrightarrow} X \mid I_t - Y_t \text{ for } k = 1, 2, \ldots, h;$$

(iii) Y does not cause X at any horizon given $I_t - Y_t$, denoted $Y \underset{(\infty)}{\nrightarrow} X \mid I_t - Y_t$ *iff*

$$Y \underset{k}{\nrightarrow} X \mid I_t - Y_t \text{ for all } k = 1, 2, \ldots$$

This definition corresponds to unidirectional causality from Y to X. It means that Y causes X at horizon h if the past of Y improves the forecast of $X(t+h)$ based on the information set $I_t - Y_t$. An alternative definition can be expressed in terms of the variance-covariance matrix of the forecast errors.

Definition 2 *For $h \geq 1$,*

(i) Y does not cause X at horizon h given $I_t - Y_t$ iff

$$\det \Sigma\big(X(t+h) \mid I_t - Y_t\big) = \det \Sigma\big(X(t+h) \mid I_t\big), \forall t > w;$$

where $\det \Sigma\big(X(t+h) \mid A_t\big)$, *represents the determinant of the variance-covariance matrix of the forecast error of $X(t+h)$ given $A_t = I_t,\ I_t - Y_t$;*

(ii) Y does not cause X up to horizon h given $I_t - Y_t$ iff $\forall\ t > w$ and $k = 1, 2, \ldots, h$,

$$\det \Sigma\big(X(t+k) \mid I_t - Y_t\big) = \det \Sigma\big(X(t+k) \mid I_t\big);$$

(iii) Y does not cause X at any horizon given $I_t - Y_t$, if $\forall\ t > w$ and $k = 1, 2, \ldots$

$$\det \Sigma\big(X(t+k) \mid I_t - Y_t\big) = \det \Sigma\big(X(t+k) \mid I_t\big).$$

[1] *The "starting point" w is not specified. In particular w may equal $-\infty$ or 0 depending on whether we consider a stationary process on the integers ($t \in \mathbb{Z}$) or a process $\{X(t) : t \geq 1\}$ on the positive integers given initial values preceding date 1.*

1.4 Causality measures

In the remainder of this chapter, we consider an information set I_t which contains two vector valued random variables of interest X and Y, and an auxiliary valued random variable Z. In other words, we suppose that $I_t = H \cup \underline{X}_t \cup \underline{Y}_t \cup \underline{Z}_t$, where H represents a subspace of the Hilbert space, possibly empty, containing time independent variables (*e.g.*, the constant in a regression model).

The causality measures we consider are extensions of the measure introduced by Geweke (1982,1984). Important properties of these measures include: 1) they are non-negative, and 2) they cancel only when there is no causality at the horizon considered. Specifically, we propose the following causality measures at horizon $h \geq 1$.

Definition 3 *For $h \geq 1$, a causality measure from Y to X at horizon h, called the* intensity *of the causality from Y to X at horizon h is given by*

$$C(Y \underset{h}{\rightarrow} X \mid Z) = \ln \left[\frac{\det \Sigma(X(t+h) \mid I_t - \underline{Y}_t)}{\det \Sigma(X(t+h) \mid I_t)} \right].$$

Remark 1 *For $m_1 = m_2 = m_3 = 1$,*

$$C(Y \underset{h}{\rightarrow} X \mid Z) = \ln \left[\frac{\sigma^2(X(t+h) \mid I_t - \underline{Y}_t)}{\sigma^2(X(t+h) \mid I_t)} \right].$$

$C(Y \underset{h}{\rightarrow} X \mid Z)$ measures the degree of the causal effect from Y to X at horizon h given the past of X and Z. In terms of predictability, this can be viewed as the amount of information brought by the past of Y that can improve the forecast of $X(t+h)$. Following Geweke (1982), this measure can be also interpreted as the proportional reduction in the variance of the forecast error of $X(t+h)$ obtained by taking into account the past of Y. This proportion is equal to:

$$\frac{\sigma^2(X(t+h) \mid I_t - \underline{Y}_t) - \sigma^2(X(t+h) \mid I_t)}{\sigma^2(X(t+h) \mid I_t - \underline{Y}_t)} = 1 - \exp[-C(Y \underset{h}{\rightarrow} X \mid Z)].$$

We can rewrite the conditional causality measures given by Definition 3 in terms of unconditional causality measures:[2]

$$C(Y \underset{h}{\rightarrow} X \mid Z) = C(YZ \underset{h}{\rightarrow} X) - C(Z \underset{h}{\rightarrow} X)$$

where

$$C(YZ \underset{h}{\rightarrow} X) = \ln\left[\frac{\det \Sigma(X(t+h) \mid I_t - \text{Y}_t - \text{Z}_t)}{\det \Sigma(X(t+h) \mid I_t)}\right],$$

$$C(Z \underset{h}{\rightarrow} X) = \ln\left[\frac{\det \Sigma(X(t+h) \mid I_t - \text{Y}_t - \text{Z}_t)}{\det \Sigma(X(t+h) \mid I_t - \text{Y}_t)}\right].$$

$C(YZ \underset{h}{\rightarrow} X)$ and $C(Z \underset{h}{\rightarrow} X)$ represent the unconditional causality measures from $(Y', Z')'$ to X and from Z to X, respectively. Similarly, we have:

$$C(X \underset{h}{\rightarrow} Y \mid Z) = C(XZ \underset{h}{\rightarrow} Y) - C(Z \underset{h}{\rightarrow} Y)$$

where

$$C(XZ \underset{h}{\rightarrow} Y) = \ln\left[\frac{\det \Sigma(Y(t+h) \mid I_t - \text{X}_t - \text{Z}_t)}{\det \Sigma(Y(t+h) \mid I_t)}\right],$$

$$C(Z \underset{h}{\rightarrow} Y) = \ln\left[\frac{\det \Sigma(Y(t+h) \mid I_t - \text{X}_t - \text{Z}_t)}{\det \Sigma(Y(t+h) \mid I_t - \text{X}_t)}\right].$$

We define an instantaneous causality measure between X and Y at horizon h as follows.

Definition 4 *For $h \geq 1$, an instantaneous causality measure between Y and X at horizon h, called the* intensity *of the instantaneous causality between Y and X at horizon h.*

[2]See Geweke (1984).

denoted $C(X \underset{h}{\leftrightarrow} Y \mid Z)$, *is given by:*

$$C(X \underset{h}{\leftrightarrow} Y \mid Z) = \ln\left[\frac{\det\Sigma(X(t+h) \mid I_t)\ \det\Sigma(Y(t+h) \mid I_t)}{\det\Sigma(X(t+h), Y(t+h) \mid I_t)}\right]$$

where $\det\Sigma(X(t+h), Y(t+h) \mid I_t)$ *represents the determinant of the variance-covariance matrix of the forecast error of the joint process* $(X', Y')'$ *at horizon* h *given the information set* I_t.

Remark 2 *For* $m_1 = m_2 = m_3 = 1$,

$$\begin{aligned}\det\Sigma\big((X(t+h), Y(t+h) \mid I_t\big) &= \sigma^2\big(X(t+h) \mid I_t\big)\,\sigma^2\big(Y(t+h) \mid I_t\big)\\ &\quad -\big(\mathsf{cov}\big((X(t+h), Y(t+h) \mid I_t\big)\big)^2. \end{aligned} \tag{1.9}$$

So the instantaneous causality measure between X *and* Y *at horizon* h *can be written as:*

$$C(X \underset{h}{\leftrightarrow} Y \mid Z) = \ln\left[\frac{1}{1-\rho^2(X(t+h), Y(t+h) \mid I_t)}\right]$$

where

$$\rho\big(X(t+h), Y(t+h) \mid I_t\big) = \frac{\mathsf{cov}\big(X(t+h), Y(t+h) \mid I_t\big)}{\sigma\big(X(t+h) \mid I_t\big)\sigma\big(Y(t+h) \mid I_t\big)}. \tag{1.10}$$

is the correlation coefficient between $X(t+h)$ *and* $Y(t+h)$ *given the information set* I_t. *Thus, the instantaneous causality measure is higher when the correlation coefficient becomes higher.*

We also define a measure of dependence between X and Y at horizon h. This will enable us to check if, at given horizon h, the processes X and Y must be considered together or whether they can be treated separately given the information set $I_t - \underline{Y}_t$.

Definition 5 *For* $h \geq 1$, *a measure of dependence between* X *and* Y *at horizon* h, *called* the intensity *of the dependence between* X *and* Y *at horizon* h, *denoted* $C^{(h)}(X, Y \mid Z)$, *is given by:*

$$C^{(h)}(X, Y \mid Z) = \ln\left[\frac{\det\Sigma(X(t+h) \mid I_t - Y_t)\ \det\Sigma(Y(t+h) \mid I_t - X_t)}{\det\Sigma(X(t+h),\ Y(t+h) \mid I_t)}\right].$$

We can easily show that the intensity of the dependence between X and Y at horizon h is equal to the sum of feedbacks measures from X to Y, from Y to X, and the instantaneous measure at horizon h. We have:

$$C^{(h)}(X, Y \mid Z) = C(X \underset{h}{\longrightarrow} Y \mid Z) + C(Y \underset{h}{\longrightarrow} X \mid Z) + C(X \underset{h}{\leftrightarrow} Y \mid Z). \tag{1.11}$$

Now, it is possible to build a recursive formulation of causality measures. This one will depend on the predictability measure introduced by Diebold and Kilian (1998). These authors proposed a predictability measure based on the ratio of expected losses of short and long run forecasts:

$$\bar{P}(L, \Omega_t, j, k) = 1 - \frac{E(L(e_{t+j,t}))}{E(L(e_{t+k,t}))}$$

where Ω_t is the information set at time t, L is a loss function, j and k represent respectively the short and the long-run, $e_{t+s,t} = X(t+s) - P(X(t+s) \mid \Omega_t)$, $s = j, k$, is the forecast error at horizon $t+s$. This predictability measure can be constructed according to the horizons of interest and it allows for general loss functions as well as univariate or multivariate information sets. In this chapter we focus on the case of a quadratic loss function,

$$L(e_{t+s,t}) = e^2_{t+s,t}, \text{ for } s = j, k.$$

We have the following relationships.

Proposition 1 *Let h_1, h_2 be two different horizons. For $h_2 > h_1 \geq 1$ and $m_1 = m_2 = 1$,*

$$C(Y \underset{h_1}{\longrightarrow} X|Z) - C(Y \underset{h_2}{\longrightarrow} X|Z) = \ln\left[1 - \bar{P}_X(I_t - Y_t, h_1, h_2)\right] - \ln\left[1 - \bar{P}_X(I_t, h_1, h_2)\right]$$

where $\bar{P}_X(\,\cdot\,, h_1, h_2)$ *represents the predictability measure for variable* X,

$$\bar{P}_X(I_t - Y_t,\, h_1,\, h_2) = 1 - \frac{\sigma^2(X(t+h_1) \mid I_t - Y_t)}{\sigma^2(X(t+h_2) \mid I_t - Y_t)},$$

$$\bar{P}_X(I_t,\, h_1,\, h_2) = 1 - \frac{\sigma^2(X(t+h_1) \mid I_t)}{\sigma^2(X(t+h_2) \mid I_t)}.$$

The following corollary follows immediately from the latter proposition.

Corollary 1 *For* $h \geq 2$ *and* $m_1 = m_2 = 1$,

$$C(Y \underset{h}{\longrightarrow} X|Z) = C(Y \underset{1}{\longrightarrow} X|Z) + \ln[1 - \bar{P}_X(I_t,\, 1,\, h)] - \ln[1 - \bar{P}_X(I_t - Y_t,\, 1,\, h)]\,.$$

For $h_2 \gg h_1$, the function $\bar{P}_k(\,.\,, h_1, h_2)$, $k = X, Y$, represents the measure of short-run forecast relative to the long-run forecast, and $C(k \underset{h_1}{\longrightarrow} l \mid Z) - C(k \underset{h_2}{\longrightarrow} l \mid Z)$, for $l \neq k$ and $l, k = X, Y$, represents the difference between the degree of short run causality and that of long run causality. Further, $\bar{P}_k(.., h_1, h_2) \gg 0$ means that the series is highly predictable at horizon h_1 relative to h_2, and $\bar{P}_k(\,.\,, h_1, h_2) = 0$, means that the series is nearly unpredictable at horizon h_1 relative to h_2.

1.5 Parametric causality measures

We now consider a more specific set of linear invertible processes which includes VAR, VMA, and VARMA models of finite order as special cases. Under this set we show that it is possible to obtain parametric expressions for short-run and long-run causality measures in terms of impulse response coefficients of a VMA representation.

This section is divided into two subsections. In the first we calculate parametric measures of short-run and long-run causality in the context of an autoregressive moving average model. We assume that the process $\{W(s) = (X'(s), Y'(s), Z'(s))' : s \leq t\}$ is a VARMA(p,q) model (hereafter unconstrained model), where p and q can be infinite. The model of the process $\{S(s) = (X'(s), Z'(s))' : s \leq t\}$ (hereafter constrained model) can be deduced from the unconstrained model using Corollary 6.1.1 in Lütkepohl (1993). This model follows a VARMA($\bar{p}$, $\bar{q}$) model with $\bar{p} \leq mp$ and $\bar{q} \leq (m-1)p+q$. In the second subsection we provide a characterization of the parametric measures in the context of VMA(q) model, where q is finite.

1.5.1 Parametric causality measures in the context of a VARMA(p, q) process

Without loss of generality, let us consider the discrete vector process with zero mean $\{W(s) = (X'(s), Y'(s), Z'(s))', s \leq t\}$ defined on L^2 and characterized by the following autoregressive moving average representation:

$$
\begin{aligned}
W(t) &= \sum_{j=1}^{p} \pi_j W(t-j) + \sum_{j=1}^{q} \varphi_j u(t-j) + u(t) \\
&= \sum_{j=1}^{p} \begin{bmatrix} \pi_{XXj} & \pi_{XYj} & \pi_{XZj} \\ \pi_{YXj} & \pi_{YYj} & \pi_{YZj} \\ \pi_{ZXj} & \pi_{ZYj} & \pi_{ZZj} \end{bmatrix} \begin{bmatrix} X(t-j) \\ Y(t-j) \\ Z(t-j) \end{bmatrix} \\
&\quad + \sum_{j=1}^{q} \begin{bmatrix} \varphi_{XXj} & \varphi_{XYj} & \varphi_{XZj} \\ \varphi_{YXj} & \varphi_{YYj} & \varphi_{YZj} \\ \varphi_{ZXj} & \varphi_{ZYj} & \varphi_{ZZj} \end{bmatrix} \begin{bmatrix} u_X(t-j) \\ u_Y(t-j) \\ u_Z(t-j) \end{bmatrix} + \begin{bmatrix} u_X(t) \\ u_Y(t) \\ u_Z(t) \end{bmatrix}
\end{aligned} \tag{1.12}
$$

with

$$
E[u(t)] = 0, \quad E\left[u(t)u'(s)\right] = \begin{cases} \Sigma_u & \text{for } s = t \\ 0 & \text{for } s \neq t \end{cases}
$$

or, more compactly,

$$
\Pi(L)W(t) = \varphi(L)u(t)
$$

where

$$\Pi(L) = \begin{bmatrix} \pi_{XX}(L) & \pi_{XY}(L) & \pi_{XZ}(L) \\ \pi_{YX}(L) & \pi_{YY}(L) & \pi_{YZ}(L) \\ \pi_{ZX}(L) & \pi_{ZY}(L) & \pi_{ZZ}(L) \end{bmatrix},$$

$$\varphi(L) = \begin{bmatrix} \varphi_{XX}(L) & \varphi_{XY}(L) & \varphi_{XZ}(L) \\ \varphi_{YX}(L) & \varphi_{YY}(L) & \varphi_{YZ}(L) \\ \varphi_{ZX}(L) & \varphi_{ZY}(L) & \varphi_{ZZ}(L) \end{bmatrix},$$

$$\pi_{ii}(L) = I_{m_i} - \sum_{j=1}^{p} \pi_{iij} L^j,\ \pi_{ik}(L) = -\sum_{j=1}^{p} \pi_{ikj} L^j,$$

$$\varphi_{ii}(L) = I_{m_i} + \sum_{j=1}^{q} \varphi_{iij} L^j,\ \ \varphi_{ik}(L) = \sum_{j=1}^{q} \varphi_{ikj} L^j,\ \text{ for } i \neq k,\ \ i,\ k = X,\ Y,\ Z.$$

We assume that $u(t)$ is orthogonal to the Hilbert subspace $\{W(s),\ s \leq (t-1)\}$ and that Σ_u is a symmetric positive definite matrix. Under stationarity, $W(t)$ has a VMA(∞) representation,

$$W(t) = \Psi(L)u(t) \tag{1.13}$$

where

$$\Psi(L) = \Pi(L)^{-1}\varphi(L) = \sum_{j=0}^{\infty} \psi_j L^j = \sum_{j=0}^{\infty} \begin{bmatrix} \psi_{XXj} & \psi_{XYj} & \psi_{XZj} \\ \psi_{YXj} & \psi_{YYj} & \psi_{YZj} \\ \psi_{ZXj} & \psi_{ZYj} & \psi_{ZZj} \end{bmatrix} L^j,\ \ \psi_0 = I_m.$$

From the previous section, measures of dependence and feedback effects are defined in terms of variance-covariance matrices of the constrained and unconstrained forecast errors. So to calculate these measures, we need to know the structure of the constrained model (imposing noncausality). This one can be deduced from the structure of the unconstrained model (1.12) using the following proposition and corollary [see Lütkepohl

(1993)].

Proposition 2 *(Linear transformation of a VMA(q) process). Let $u(t)$ be a K-dimensional white noise process with nonsingular variance-covariance matrix Σ_u and let*

$$W(t) = \mu + \sum_{j=1}^{q} \Psi_j u(t-j) + u(t),$$

be a K-dimensional invertible VMA(q) process. Furthermore, let F be an $(M \times K)$ matrix of rank M. Then the M-dimensional process $S(t) = FW(t)$ has an invertible VMA($\bar{q}$) representation:

$$S(t) = F\mu + \sum_{j=1}^{\bar{q}} \theta_j \varepsilon(t-j) + \varepsilon(t),$$

where $\varepsilon(t)$ is M-dimensional white noise with nonsingular variance-covariance matrix Σ_ε, the θ_j, $j = 1, \ldots, \bar{q}$, are $(M \times M)$ coefficient matrices and $\bar{q} \leq q$.

Corollary 2 *(Linear Transformation of a VARMA(p,q) process). Let $W(t)$ be a K-dimensional, stable, invertible VARMA(p,q) process and let F be an $(M \times K)$ matrix of rank M. Then the process $S(t) = FW(t)$ has a VARMA($\bar{p}$, $\bar{q}$) representation with*

$$\bar{p} \leq Kp, \quad \bar{q} \leq (K-1)p + q.$$

Remark 3 *If we assume that $W(t)$ follows a VAR(p)$\equiv$VARMA(p,0) model, then its linear transformation $S(t) = FW(t)$ has a VARMA($\bar{p}$, $\bar{q}$) representation with $\bar{p} \leq Kp$ and $\bar{q} \leq (K-1)p$.*

Suppose that we are interested in measuring the causality from Y to X at a given horizon h. We need to apply Corollary (2) to define the structure of process $S(s) = (X(s)', Z(s)')'$. If we left-multiply equation (1.13) by the adjoint matrix of $\Pi(L)$, denoted $\Pi(L)^*$, we get

$$\Pi(L)^* \Pi(L) W(t) = \Pi(L)^* \varphi(L) u(t) \tag{1.14}$$

where $\Pi(L)^*\Pi(L) = \det[\Pi(L)]$. Since the determinant of $\Pi(L)$ is a sum of products involving one operator from each row and each column of $\Pi(L)$, the degree of the AR polynomial, here $\det[\Pi(L)]$, is at most mp. We write:

$$\det[\Pi(L)] = 1 - \alpha_1 L - \cdots - \alpha_{\bar{p}} L^{\bar{p}}$$

where $\bar{p} \leq mp$. It is also easy to check that the degree of the operator $\Pi(L)^*\varphi(L)$ is at most $p(m-1)+q$. Thus, Equation (1.14) can be written as follows:

$$\det[\Pi(L)]\, W(t) = \Pi(L)^*\varphi(L)u(t). \tag{1.15}$$

This equation is another stationary invertible VARMA representation of process $W(t)$, called the final equation form. The model of process $\{S(s) = (X'(s), Z'(s))', s \leq t\}$ can be obtained by choosing

$$F = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & 0 & I_{m3} \end{bmatrix}.$$

On premultiplying (1.15) by F, we get

$$\det[\Pi(L)]\, S(t) = F\Pi(L)^*\varphi(L)u(t). \tag{1.16}$$

The right-hand side of Equation (1.16) is a linearly transformed finite order VMA process which, by Proposition 2, has a VMA($\bar{q}$) representation with $\bar{q} \leq p(m-1)+q$. Thus, we get the following constrained model:

$$\det[\Pi(L)]\, S(t) = \theta(L)\varepsilon(t) = \begin{bmatrix} \theta_{XX}(L) & \theta_{XZ}(L) \\ \theta_{ZX}(L) & \theta_{ZZ}(L) \end{bmatrix} \varepsilon(t) \tag{1.17}$$

where

$$E[\varepsilon(t)] = 0,\ E\left[\varepsilon(t)\varepsilon'(s)\right] = \begin{cases} \Sigma_\varepsilon & \text{for } s = t \\ 0 & \text{for } s \neq t \end{cases}.$$

$$\theta_{ii}(L) = I_{m_i} + \sum_{j=1}^{\bar{q}} \theta_{ii,j} L^j, \; \theta_{ik}(L) = \sum_{j=1}^{\bar{q}} \theta_{ikj} L^j, \quad \text{for } i \neq k, \; i, k = X, Z.$$

Note that, in theory, the coefficients θ_{ikj}, $i, k = X, Z$; $j = 1, ...\bar{q}$, and elements of the variance-covariance matrix Σ_ε, can be computed from coefficients π_{ikj} , φ_{ikl}, $i, k = X$, Z, Y; $j = 1, \ldots, p$; $l = 1, \ldots, q$, and elements of the variance-covariance matrix Σ_u. This is possible by solving the following system:

$$\gamma_\varepsilon(v) = \gamma_u(v), \quad v = 0, 1, 2, \ldots \tag{1.18}$$

where $\gamma_\varepsilon(v)$ and $\gamma_u(v)$ are the autocovariance functions of the processes $\theta(L)\varepsilon(t)$ and $F\Pi(L)^*\varphi(L)u(t)$, respectively. For large numbers m, p, and q, system (1.18) can be solved by using optimization methods.[3] The following example shows how one can calculate the theoretical parameters of the constrained model in terms of those of the unconstrained model in the context of a bivariate VAR(1) model.

Example 3 *Consider the following bivariate VAR(1) model:*

$$\begin{aligned} \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} &= \begin{bmatrix} \pi_{XX} & \pi_{XY} \\ \pi_{YX} & \pi_{YY} \end{bmatrix} \begin{bmatrix} X(t-1) \\ Y(t-1) \end{bmatrix} + \begin{bmatrix} u_X(t) \\ u_Y(t) \end{bmatrix} \\ &= \pi \begin{bmatrix} X(t-1) \\ Y(t-1) \end{bmatrix} + u(t). \end{aligned} \tag{1.19}$$

We assume that all roots of $\det[\Pi(z)] = \det(I_2 - \pi z)$ *are outside of the unit circle. Under this assumption model (1.19) has the following VMA(∞) representation:*

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \sum_{j=0}^{\infty} \psi_j \begin{pmatrix} u_X(t-j) \\ u_Y(t-j) \end{pmatrix} = \sum_{j=0}^{\infty} \begin{bmatrix} \psi_{XX,j} & \psi_{XY,j} \\ \psi_{YX,j} & \psi_{YY,j} \end{bmatrix} \begin{pmatrix} u_X(t-j) \\ u_Y(t-j) \end{pmatrix}$$

[3]In section 1.7 we discuss another approach to computing the constrained model using simulation technique.

where

$$\psi_j = \pi\psi_{j-1} = \pi^j, \quad j = 1, 2, \dots, \quad \psi_0 = I_2.$$

If we are interested in determining the model of marginal process $X(t)$, *then by Corollary (2) and for* $F = [1, 0]$, *we have*

$$\det[\Pi(L)]X(t) = [1, 0]\,\Pi(L)^* u(t)$$

where

$$\Pi(L)^* = \begin{bmatrix} 1 - \pi_{YY}L & \pi_{XY}L \\ \pi_{YX}L & 1 - \pi_{XX}L \end{bmatrix},$$

and

$$\det[\Pi(L)] = 1 - (\pi_{YY} + \pi_{XX})L - (\pi_{YX}\pi_{XY} - \pi_{XX}\pi_{YY})L^2. \tag{1.20}$$

Thus,

$$X(t) - \pi_1 X(t-1) - \pi_2 X(t-2) = \pi_{XY}u_Y(t-1) - \pi_{YY}u_X(t-1) + u_X(t).$$

where $\pi_1 = \pi_{YY} + \pi_{XX}$ *and* $\pi_2 = \pi_{YX}\pi_{XY} - \pi_{XX}\pi_{YY}$. *The right-hand side of equation 1.20, denoted* $\varpi(t)$, *is the sum of an MA(1) process and a white noise process. By Proposition 2,* $\varpi(t)$ *has an MA(1) representation,* $\varpi(t) = \varepsilon_X(t) + \theta\varepsilon_X(t-1)$. *To determine parameters* θ *and* $Var(\varepsilon_X(t)) = \sigma^2_{\varepsilon_X}$ *in terms of the parameters of the unconstrained model, we have to solve system (1.18) for* $v = 0$ *and* $v = 1$,

$$Var\left[\varpi(t)\right] = Var\left[u_X(t) - \pi_{YY}u_X(t-1) + \pi_{XY}u_Y(t-1)\right],$$

$$E\left[\varpi(t)\varpi(t-1)\right] = E[(u_X(t) - \pi_{YY}u_X(t-1) + \pi_{XY}u_Y(t-1))$$

$$\times\,(u_X(t-1) - \pi_{YY}u_X(t-2) + \pi_{XY}u_Y(t-2))].$$

$\Longleftrightarrow$

$$(1+\theta^2)\sigma^2_{\varepsilon_X} = (1+\pi^2_{YY})\sigma^2_{u_X} + \pi^2_{XY}\sigma^2_{u_Y} - 2\pi_{YY}\pi_{XY}\sigma_{u_Y u_X}.$$

$$\theta\sigma^2_{\varepsilon_X} = -\pi_{YY}\sigma^2_{u_X}.$$

Here we have two equations and two unknown parameters θ *and* $\sigma^2_{\varepsilon_X}$. *These parameters must satisfy the constraints* $\mid \theta \mid < 1$ *and* $\sigma^2_{\varepsilon_X} > 0$.

The VMA(∞) representation of model (1.17) is given by:

$$\begin{aligned} S(t) &= \det[\Pi(L)]^{-1}\theta(L)\varepsilon(t) = \sum_{j=0}^{\infty}\Phi_j\varepsilon(t-j) \\ &= \sum_{j=0}^{\infty}\begin{bmatrix} \phi_{XX_j} & \phi_{XZ_j} \\ \phi_{ZX_j} & \phi_{ZZ_j} \end{bmatrix}\begin{bmatrix} \varepsilon_X(t-j) \\ \varepsilon_Z(t-j) \end{bmatrix}. \end{aligned} \tag{1.21}$$

where $\Phi_0 = I_{(m_1+m_2)}$. To quantify the degree of causality from Y to X at horizon h, we first consider the unconstrained and constrained models of process X. The unconstrained model is given by the following equation:

$$X(t) = \sum_{j=1}^{\infty}\psi_{XX_j}u_X(t-j) + \sum_{j=1}^{\infty}\psi_{XY_j}u_Y(t-j) + \sum_{j=1}^{\infty}\psi_{XZ_j}u_Z(t-j) + u_X(t),$$

whereas the constrained model is given by:

$$X(t) = \sum_{j=1}^{\infty}\phi_{XX_j}\varepsilon_X(t-j) + \sum_{j=1}^{\infty}\phi_{XZ_j}\varepsilon_Z(t-j) + \varepsilon_X(t).$$

Second, we need to calculate the variance-covariance matrices of the unconstrained and constrained forecast errors of $X(t+h)$. From equation (1.13), the forecast error of $W(t+h)$ is given by:

$$e_{nc}[W(t+h) \mid I_t] = \sum_{i=0}^{h-1}\psi_i u(t+h-i),$$

associated with the variance-covariance matrix

$$\Sigma(W(t+h) \mid I_t) = \sum_{i=0}^{h-1} \psi_i \; Var\left[u(t)\right] \psi_i' = \sum_{i=0}^{h-1} \psi_i \; \Sigma_u \psi_i'. \tag{1.22}$$

The unconstrained forecast error of $X(t+h)$ is given by

$$\begin{aligned} e_{nc}\left[X(t+h) \mid I_t\right] &= \sum_{j=1}^{h-1} \psi_{XXj} u_X(t+h-j) + \sum_{j=1}^{h-1} \psi_{XYj} u_Y(t+h-j) \\ &\quad + \sum_{j=1}^{h-1} \psi_{XZj} u_Z(t+h-j) + u_X(t+h), \end{aligned}$$

which is associated with the unconstrained variance-covariance matrix

$$\Sigma(X(t+h) \mid I_t) = \sum_{i=0}^{h-1} [e_X^{nc} \psi_i \Sigma_u \psi_i' e_X^{nc\prime}],$$

where $e_X^{nc} = \begin{bmatrix} I_{m_1} & 0 & 0 \end{bmatrix}$. Similarly, the forecast error of $S(t+h)$ is given by

$$e_c\left[S(t+h) \mid I_t - \mathrm{Y}_t\right] = \sum_{i=0}^{h-1} \Phi_i \varepsilon(t+h-i)$$

associated with the variance-covariance matrix

$$\Sigma(S(t+h) \mid I_t - \mathrm{Y}_t) = \sum_{i=0}^{h-1} \Phi_i \; \Sigma_\varepsilon \Phi_i' \,.$$

Consequently, the constrained forecast error of $X(t+h)$ is given by:

$$e_c\left[X(t+h) \mid I_t - \mathrm{Y}_t\right] = \sum_{j=1}^{h-1} \phi_{XXj} \; \varepsilon_X(t+h-j) + \sum_{j=1}^{h-1} \phi_{XZj} \; \varepsilon_Z(t+h-j) + \varepsilon_X(t+h)$$

associated with the constrained variance-covariance matrix

$$\Sigma(X(t+h) \mid I_t - Y_t) = \sum_{i=0}^{h-1} e_X^c \Phi_i \, \Sigma_\varepsilon \Phi_i' \, e_X^{c'}$$

where $e_X^c = \begin{bmatrix} I_{m_1} & 0 \end{bmatrix}$. Thus, we can immediately deduce the following result by using the definition of a causality measure from Y to X [see Definition 3].

Theorem 3 *Under assumptions (1.12) and (1.13): and for $h \geq 1$, where h is a positive integer,*

$$C(Y \underset{h}{\longrightarrow} X|Z) = \ln\left[\frac{\det(\sum_{i=0}^{h-1} e_X^c \Phi_i \, \Sigma_\varepsilon \Phi_i' \, e_X^{c'})}{\det(\sum_{i=0}^{h-1} e_X^{nc} \psi_i \Sigma_u \psi_i' e_X^{nc'})}\right],$$

where $e_X^{nc} = \begin{bmatrix} I_{m_1} & 0 & 0 \end{bmatrix}$, $e_X^c = \begin{bmatrix} I_{m_1} & 0 \end{bmatrix}$.

We can, of course, repeat the same argument switching the role of the variables X and Y.

Example 4 *For a bivariate VAR(1) model [see Example 3], we can analytically compute the causality measures at any horizon h using only the unconstrained parameters. For example, the causality measures at horizons 1 and 2 are given by:*[4]

$$C(Y \underset{1}{\longrightarrow} X) = \ln\left[\frac{(1+\pi_{YY}^2)\sigma_{u_X}^2 + \pi_{XY}^2 \sigma_{u_Y}^2 + \sqrt{((1+\pi_{YY}^2)\sigma_{u_X}^2 + \pi_{XY}^2 \sigma_{u_Y}^2)^2 - 4\pi_{YY}^2 \sigma_{u_X}^4}}{2\sigma_{u_X}^2}\right], \tag{1.23}$$

$$C(Y \underset{2}{\longrightarrow} X) = \tag{1.24}$$

$$\ln\left[\frac{4\pi_{YY}^2 \sigma_{u_X}^4 + [(1+\pi_{YY}^2)\sigma_{u_X}^2 + \pi_{XY}^2 \sigma_{u_Y}^2 - \sqrt{((1+\pi_{YY})^2 \sigma_{u_X}^2 + \pi_{XY}^2 \sigma_{u_Y}^2)^2 - 4\pi_{YY}^2 \sigma_{u_X}^4} - 2\pi_{YY}\sigma_{u_X}^2]^2}{2[(1+\pi_{XX}^2)\sigma_{u_X}^2 + \pi_{XY}^2 \sigma_{u_Y}^2][(1+\pi_{YY}^2)\sigma_{u_X}^2 + \pi_{XY}^2 \sigma_{u_Y}^2 - \sqrt{((1+\pi_{YY})^2 \sigma_{u_X}^2 + \pi_{XY}^2 \sigma_{u_Y}^2)^2 - 4\pi_{YY}^2 \sigma_{u_X}^4}]}\right].$$

[4] *Equations 1.23 and 1.24 are obtained under assumptions $cov(u_X(t), u_Y(t)) = 0$ and*

$$((1+\pi_{YY}^2)\sigma_{u_X}^2 + \pi_{XY}^2 \sigma_{u_Y}^2)^2 - 4\pi_{YY}^2 \sigma_{u_X}^4 \geq 0.$$

Now we will determine the parametric measure of instantaneous causality at given horizon h. We know from section 1.4 that a measure of instantaneous causality is defined only in terms of the variance-covariance matrices of unconstrained forecast errors. The variance-covariance matrix of the unconstrained forecast error of joint process $(X'(t+h), Y'(t+h))'$ is given by:

$$\Sigma(X(t+h), Y(t+h) \mid I_t) = \sum_{i=0}^{h-1} G\psi_i \Sigma_u \psi_i' G'.$$

where $G = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & I_{m_2} & 0 \end{bmatrix}$. We have,

$$\Sigma(X(t+h) \mid I_t) = \sum_{i=0}^{h-1} [e_X^{nc} \psi_i \Sigma_u \psi_i' e_X^{nc'}],$$

$$\Sigma(Y(t+h) \mid I_t) = \sum_{i=0}^{h-1} [e_Y^{nc} \psi_i \Sigma_u \psi_i' e_Y^{nc'}],$$

where $e_Y^{nc} = \begin{bmatrix} 0 & I_{m2} & 0 \end{bmatrix}$. Thus, we can immediately deduce the following result by using the definition of the instantaneous causality measure.

Theorem 4 *Under assumptions (1.12) and (1.13) and for $h \geq 1$.*

$$C(X \underset{h}{\longleftrightarrow} Y|Z) = \ln \left[\frac{\det(\sum_{i=0}^{h-1} [e_X^{nc} \psi_i \Sigma_u \psi_i' e_X^{nc'}]) \ \det(\sum_{i=0}^{h-1} [e_Y^{nc} \psi_i \Sigma_u \psi_i' e_Y^{nc'}])}{\det(\sum_{i=0}^{h-1} [G \psi_i \Sigma_u \psi_i' G'])} \right]$$

where $G = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & I_{m_2} & 0 \end{bmatrix}$, $e_X = \begin{bmatrix} I_{m_1} & 0 & 0 \end{bmatrix}$, $e_Y = \begin{bmatrix} 0 & I_{m_2} & 0 \end{bmatrix}$.

The parametric measure of dependence at horizon h can be deduced from its decomposition given by equation (1.11).

1.5.2 Characterization of causality measures for VMA(q) processes

Now, assume that the process $\{W(s) = (X'(s), Z'(s), Y'(s))' : s \leq t\}$ follows an invertible VMA(q) model:

$$\begin{aligned} W(t) &= \sum_{j=1}^{q} \Phi_j u(t-j) + u(t) \\ &= \sum_{j=1}^{q} \begin{bmatrix} \Phi_{XXj} & \Phi_{XYj} & \Phi_{XZj} \\ \Phi_{YXj} & \Phi_{YYj} & \Phi_{YZj} \\ \Phi_{ZXj} & \Phi_{ZYj} & \Phi_{ZZj} \end{bmatrix} \begin{bmatrix} u_X(t-j) \\ u_Y(t-j) \\ u_Z(t-j) \end{bmatrix} + \begin{bmatrix} u_X(t) \\ u_Y(t) \\ u_Z(t) \end{bmatrix}. \end{aligned} \quad (1.25)$$

More compactly,

$$W(t) = \Phi(L)u(t)$$

where

$$\Phi(\mathrm{L}) = \begin{bmatrix} \Phi_{XX}(L) & \Phi_{XY}(L) & \Phi_{XZ}(L) \\ \Phi_{YX}(L) & \Phi_{YY}(L) & \Phi_{YZ}(L) \\ \Phi_{ZX}(L) & \Phi_{ZY}(L) & \Phi_{ZZ}(L) \end{bmatrix}.$$

$$\Phi_{ii}(L) = I_{m_i} + \sum_{j=1}^{q} \Phi_{iij} L^j,\ \Phi_{ik}(L) = \sum_{j=1}^{q} \Phi_{ikj} L^j \text{, for } i \neq k,\ \ i, k = X,\ Z,\ Y.$$

From Proposition 2 and letting $F = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix}$, the model of the constrained process $S(t) = FW(t)$ is an MA($\bar{q}$) with $\bar{q} \leq q$. We have,

$$S(t) = \theta(L)\varepsilon(t) = \sum_{j=0}^{\bar{q}} \theta_j \varepsilon(t-j) = \sum_{j=0}^{\bar{q}} \begin{bmatrix} \theta_{XX,j} & \theta_{XZ,j} \\ \theta_{ZX,j} & \theta_{ZZ,j} \end{bmatrix} \begin{pmatrix} \varepsilon_X(t-j) \\ \varepsilon_Z(t-j) \end{pmatrix}$$

where

$$E[\varepsilon(t)] = 0,\ E\left[\varepsilon(t)\varepsilon'(s)\right] = \begin{cases} \Sigma_\varepsilon & \text{for } s = t \\ 0 & \text{for } s \neq t \end{cases}.$$

Theorem 5 *Let h_1 and h_2 be two different horizons. Under assumption (1.25) we have,*

$$C(Y \underset{h_1}{\longrightarrow} X \mid Z) = C(Y \underset{h_2}{\longrightarrow} X \mid Z), \ \forall \ h_2 \geq h_1 \geq q.$$

This result follows immediately from Proposition 1.

1.6 Estimation

We know from section 1.5 that short- and long-run causality measures depend on the parameters of the model describing the process of interest. Consequently, these measures can be estimated by replacing the unknown parameters by their estimates from a finite sample.

Three different approaches for estimating causality measures can be considered. The first, called the nonparametric approach, is the focus of this section. It assumes that the form of the parametric model appropriate for the process of interest is unknown and approximates it with a VAR(k) model, where k depends on the sample size [see Parzen (1974), Bhansali (1978), and Lewis and Reinsel (1985)]. The second approach assumes that the process follows a finite order VARMA model. The standard methods for the estimation of VARMA models, such as maximum likelihood and nonlinear least squares, require nonlinear optimization. This might not be feasible because the number of parameters can increase quickly. To circumvent this problem, several authors [see Hannan and Rissanen (1982), Hannan and Kavalieris (1984b), Koreisha and Pukkila (1989), Dufour and Pelletier (2005), and Dufour and Jouini (2004)] have developed a relatively simple approach based only on linear regression. This approach enables estimation of VARMA models using a long VAR whose order depends on the sample size. The last and simplest approach assumes that the process follows a finite order VAR(p) model which can be estimated by OLS.

In practice, the precise form of the parametric model appropriate for a process is unknown. Parzen (1974), Bhansali (1978), and Lewis and Reinsel (1985), among others, considered a nonparametric approach to predicting future values using an autoregressive model fitted to a series of T observations. This approach is based on a very mild assumption of an infinite order autoregressive model for the process which includes finite-order stationary VARMA processes as a special case. In this section, we describe the nonparametric approach to estimating the short- and long-run causality measures. First, we discuss estimation of the fitted autoregressive constrained and unconstrained models. We then point out some assumptions necessary for the convergence of the estimated parameters. Second, using Theorem 6 in Lewis and Reinsel (1985), we define approximations of variance-covariance matrices of the constrained and unconstrained forecast errors at horizon h. Finally, we use these approximations to construct an asymptotic estimator of short- and long-run causality measures.

In what follows we focus on the estimation of the unconstrained model. Let us consider a stationary vector process $\{W(s) = (X(s)', Y(s)', Z(s)')', s \leq t)\}$. By Wold's theorem, this process can be written in the form of a VMA(∞) model:

$$W(t) = u(t) + \sum_{j=1}^{\infty} \varphi_j u(t-j).$$

We assume that $\sum_{j=0}^{\infty} \| \varphi_j \| < \infty$ and $\det\{\varphi(z)\} \neq 0$ for $| z | \leq 1$, where $\| \varphi_j \| = tr(\varphi_j' \varphi_j)$ and $\varphi(z) = \sum_{j=0}^{\infty} \varphi_j z^j$, with $\varphi_0 = I_m$, an $m \times m$ identity matrix. Under the latter assumptions, $W(t)$ is invertible and can be written as an infinite autoregressive process:

$$W(t) = \sum_{j=1}^{\infty} \pi_j W(t-j) + u(t), \tag{1.26}$$

where $\sum_{j=1}^{\infty} \| \pi_j \| < \infty$ and $\pi(z) = I_m - \sum_{j=1}^{\infty} \pi_j z^j = \varphi(z)^{-1}$ satisfies $\det\{\pi(z)\} \neq 0$ for $| z | \leq 1$.

Let $\Pi(k) = (\pi_1, \pi_2, \ldots, \pi_k)$ denote the first k autoregressive coefficients in the VAR(∞) representation. Given a realization $\{W(1), \ldots, W(T)\}$, we can approximate (1.26) by a finite order VAR(k) model, where k depends on the sample size T. The estimators of the autoregressive coefficients of the fitted VAR(k) model and variance-covariance matrix, $\hat{\Sigma}_u^k$, are given by the following equation:

$$\hat{\Pi}(k) = (\hat{\pi}_{1k}, \hat{\pi}_{2k}, \ldots, \hat{\pi}_{kk}) = \hat{\Gamma}'_{k1}\hat{\Gamma}_k^{-1}, \quad \hat{\Sigma}_u^k = \sum_{t=k+1}^{T} \hat{u}_k(t)\hat{u}_k(t)'/(T-k),$$

where $\hat{\Gamma}_k = (T-k)^{-1}\sum_{t=k+1}^{T} w(t)w(t)'$, for $w(t) = (W(t)', \ldots, W(t-k+1)')'$, $\hat{\Gamma}_{k1} = (T-k)^{-1}\sum_{t=k+1}^{T} w(t)W(t+1)'$, and $\hat{u}_k(t) = W(t) - \sum_{j=1}^{k} \hat{\pi}_{jk} W(t-j)$.

Theorem 1 in Lewis and Reinsel (1985) ensures convergence of $\hat{\Pi}(k)$ under three assumptions: (1) $E \mid u_i(t)u_j(t)u_k(t)u_l(t) \mid \leq \gamma_4 < \infty$, for $1 \leq i, j, k, l \leq m$; (2) k is chosen as a function of T such that $k^2/T \rightarrow 0$ as $k, T \rightarrow \infty$; and (3) k is chosen as a function of T such that $k^{1/2} \sum_{j=k+1}^{\infty} \parallel \pi_j \parallel \rightarrow \infty$ as $k, T \rightarrow \infty$. In their Theorem 4 they derive the asymptotic distribution for these estimators under 3 assumptions: (1) $E \mid u_i(t)u_j(t)u_k(t)u_l(t) \mid \leq \gamma_4 < \infty, 1 \leq i, j, k, l \leq m$; (2) k is chosen as a function of T such that $k^3/T \rightarrow 0$ as $k, T \rightarrow \infty$; and (3) there exists $\{l(k)\}$ a sequence of $(km^2 \times 1)$ vectors such that $0 < M_1 \leq \parallel l(k) \parallel^2 = l(k)'l(k) \leq M_2 < \infty$, for $k = 1, 2, \ldots$ We also note that $\hat{\Sigma}_u^k$ converges to Σ_u, as k and $T \rightarrow \infty$ [see Lütkepohl (1993a)].

Remark 4 *The upper bound K of the order k in the fitted VAR(k) model depends on the assumptions required to ensure convergence and the asymptotic distribution of the estimator. For convergence of the estimator, we need to assume that $k^2/T \rightarrow 0$, as k and $T \rightarrow \infty$. Consequently, we can choose $K = CT^{1/2}$, where C is a constant, as an upper bound. To derive the asymptotic distribution of the estimator $\hat{\Pi}(k)$, we need to assume that $k^3/T \rightarrow 0$, as k and $T \rightarrow \infty$, and thus we can choose as an upper bound $K = CT^{1/3}$.*

The forecast error of $W(t+h)$, based on the $VAR(\infty)$ model, is given by:

$$e_{nc}[W(t+h) \mid W(t), W(t-1), \ldots] = \sum_{j=0}^{h-1} \varphi_j u(t+h-j).$$

associated with the variance-covariance matrix

$$\Sigma[W(t+h) \mid W(t),\, W(t-1),\, \ldots] = \sum_{j=0}^{h-1} \varphi_j \Sigma_u \varphi_j'.$$

In the same way, the variance-covariance matrix of the forecast error of $W(t+h)$, based on the VAR(k) model, is given by:

$$\Sigma_k[W(t+h) \mid W(t),\, W(t-1),\, \ldots,\, W(t-k+1)]$$

$$= E\left[\left(W(t+h) - \sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t+1-j)\right)\left(W(t+h) - \sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t+1-j)\right)'\right],$$

where [see Dufour and Renault (1998)]

$$\hat{\pi}_{jk}^{(h+1)} = \hat{\pi}_{(j+1)k}^{(h+1)} + \hat{\pi}_{1k}^{(h)} \hat{\pi}_{jk}, \quad \hat{\pi}_{jk}^{(1)} = \hat{\pi}_{jk}, \quad \hat{\pi}_{jk}^{(0)} = I_m, \quad \text{for } j \geq 1, \;\; h \geq 1;$$

Moreover,

$$W(t+h) - \sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t+1-j)$$

$$= \left(W(t+h) - \sum_{j=1}^{\infty} \pi_j^{(h)} W(t+1-j)\right) - \left(\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t+1-j) - \sum_{j=1}^{\infty} \pi_j^{(h)} W(t+1-j)\right)$$

$$= \sum_{j=0}^{h-1} \varphi_j (t+1-j) - \left(\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t+1-j) - \sum_{j=1}^{\infty} \pi_j^{(h)} W(t+1-j)\right). \tag{1.27}$$

where [see Dufour and Renault (1998)]

$$\pi_j^{(h+1)} = \pi_{j+1}^{(h+1)} + \pi_1^{(h)} \pi_j, \quad \pi_j^{(1)} = \pi_j, \quad \pi_j^{(0)} = I_m, \quad \text{for } j \geq 1 \text{ and } h \geq 1.$$

Since the error terms $u(t+h-j)$, for $0 \le j \le (h-1)$, are independent of $(W(t), W(t-1), \dots)$ and $(\hat{\pi}_{1k}, \hat{\pi}_{2k}, \dots, \hat{\pi}_{kk})$, the two terms on the right-hand side of equation (1.27) are independent. Thus,

$$\begin{aligned}
&\Sigma_k[W(t+h) \mid W(t), W(t-1), \dots, W(t-k+1)] \\
&= E[\left(\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t+1-j) - \sum_{j=1}^{\infty} \pi_j^{(h)} W(t+1-j)\right) \\
&\times \left(\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t+1-j) - \sum_{j=1}^{\infty} \pi_j^{(h)} W(t+1-j)\right)'] \\
&+\Sigma[W(t+h) \mid W(t), W(t-1), \dots].
\end{aligned} \tag{1.28}$$

As k and $T \to \infty$, an asymptotic approximation of the first term in equation (1.28) is given by Theorem 6 in Lewis and Reinsel (1985):

$$E\left[\left(\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t+1-j) - \sum_{j=1}^{\infty} \pi_j^{(h)} W(t+1-j)\right)\left(\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t+1-j) - \sum_{j=1}^{\infty} \pi_j^{(h)} W(t+1-j)\right)'\right]$$
$$\approx \frac{km}{T} \sum_{j=0}^{h-1} \varphi_j \Sigma_u \varphi_j'.$$

Consequently, an asymptotic approximation of the variance-covariance matrix of the forecast error is given by:

$$\Sigma_k[W(t+h) \mid W(t), W(t-1), \dots, W(t-k+1)] \approx (1 + \frac{km}{T}) \sum_{j=0}^{h-1} \varphi_j \Sigma_u \varphi_j'. \tag{1.29}$$

An estimator of this quantity is obtained by replacing the parameters φ_j and Σ_u by their estimators $\hat{\varphi}_j^k$ and $\hat{\Sigma}_u^k$, respectively.

We can also obtain an asymptotic approximation of the variance-covariance matrix of the constrained forecast error at horizon h following the same steps as before. We denote this variance-covariance matrix by:

$$\Sigma_k[S(t+h) \mid S(t), S(t-1), \ldots, S(t-k+1)] \approx (1 + \frac{k(m_1+m_3)}{T}) \sum_{j=0}^{h-1} \Phi_j \Sigma_\varepsilon \Phi_j',$$

where Φ_j, for $j = 1, \ldots, h-1$, represent the coefficients of a VMA representation of the constrained process S, and Σ_ε is the variance-covariance matrix of $\varepsilon(t) = (\varepsilon_X(t)', \varepsilon_Z(t)')'$. From the above results, an asymptotic approximation of the causality measure from Y to X is given by:

$$C^a(Y \underset{h}{\longrightarrow} X \mid Z) = \ln \left[\frac{\det[\sum_{j=0}^{h-1} e_X^c \Phi_j \Sigma_\varepsilon \Phi_j' e_X^{c\prime}]}{\det[\sum_{j=0}^{h-1} e_X^{nc} \varphi_j \Sigma_u \varphi_j' e_X^{nc\prime}]}\right] + \ln \left[1 - \frac{km_2}{T+km}\right]$$

where $e_X^{nc} = \left[\begin{array}{ccc} I_{m_1} & 0 & 0 \end{array}\right]$ and $e_X^c = \left[\begin{array}{cc} I_{m_1} & 0 \end{array}\right]$. An estimator of this quantity will be obtained by replacing the unknown parameters, Φ_j, Σ_ε, φ_j, and Σ_u, by their estimates, $\hat{\Phi}_j^k$, $\hat{\Sigma}_\varepsilon^k$, $\hat{\varphi}_j^k$, and $\hat{\Sigma}_u^k$, respectively:

$$\hat{C}^a(Y \underset{h}{\longrightarrow} X \mid Z) = \ln \left[\frac{\det[\sum_{j=0}^{h-1} e_X^c \hat{\Phi}_j^k \hat{\Sigma}_\varepsilon^k \hat{\Phi}_j^{k\prime} e_X^{c\prime}]}{\det[\sum_{j=0}^{h-1} e_X^{nc} \hat{\varphi}_j^k \hat{\Sigma}_u^k \hat{\varphi}_j^{k\prime} e_X^{nc\prime}]}\right] + \ln \left[1 - \frac{km_2}{T+km}\right].$$

1.7 Evaluation by simulation of causality measures

In this section, we propose a simple simulation-based technique to calculate causality measures at any horizon h, for $h \geq 1$. To illustrate this technique we consider the same examples we used in section 1 and limit ourselves to horizons 1 and 2.

Since one source of bias in autoregressive coefficients is sample size, our technique consists of simulating a large sample from the unconstrained model whose parameters are assumed to be either known or estimated from a real data set. Once the large sample (hereafter large simulation) is simulated, we use it to estimate the parameters of the constrained model (imposing noncausality). In what follows we describe an algorithm to calculate the causality measure at given horizon h using a large simulation technique:

1. given the parameters of the unconstrained model and its initial values, simulate a large sample of T observations under the assumption that the probability distribution of the error term $u(t)$ is completely specified;[5][6]

2. estimate the constrained model using a large simulation;

3. calculate the constrained and unconstrained variance-covariance matrices of the forecast errors at horizon h [see section 1.5];

4. calculate the causality measure at horizon h using the constrained and unconstrained variance-covariance matrices from step 3.

Now, let us reconsider Example 1 from section 1:

$$\begin{aligned} \begin{bmatrix} X(t+1) \\ Y(t+1) \end{bmatrix} &= \Pi \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} + u(t) \\ &= \begin{bmatrix} 0.5 & 0.7 \\ 0.4 & 0.35 \end{bmatrix} \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} + \begin{bmatrix} u_X(t+1) \\ u_Y(t+1) \end{bmatrix}, \end{aligned} \tag{1.30}$$

where

$$E[u(t)] = 0, \; E[u(t)u(s)'] = \begin{cases} I_2 \text{ if } s = t \\ 0 \text{ if } s \neq t \end{cases}.$$

Our illustration involves two steps. First, we calculate the theoretical values of the causality measures at horizons 1 and 2. We know from Example 4 that for a bivariate VAR(1) model it is easy to compute the causality measure at any horizon h using only the unconstrained parameters. Second, we evaluate the causality measures using a large simulation technique and we compare them with theoretical values from step 1. These theoretical values are recovered as follows.

1. We compute the variances of the forecast errors of X at horizons 1 and 2 using its

[5] T can be equal to 1000000, ...
[6] The form of the probability distribution of $u(t)$ does not affect the value of causality measures.

own past and the past of Y. We have,

$$\Sigma[(X(t+h),\, Y(t+h))' \mid \underline{X}_t,\, \underline{Y}_t] = \sum_{i=0}^{h-1} \Pi^i\, \Pi^{i'}. \tag{1.31}$$

From (1.31), we get

$$Var[X(t+1) \mid \underline{X}_t,\, \underline{Y}_t] = 1,\, Var[X(t+2) \mid \underline{X}_t,\, \underline{Y}_t] = \sum_{i=0}^{1} e\, \Pi^i\, \Pi^{i'} e' = 1.74,$$

where $e = (1, 0)'$.

2. We compute the variances of the forecast errors of X at horizons 1 and 2 using only its own past. In this case we need to determine the structure of the constrained model. This one is given by the following equation [see Example 3]:

$$X(t+1) = (\pi_{YY}+\pi_{XX})X(t) + (\pi_{YX}\pi_{XY}-\pi_{XX}\pi_{YY})X(t-1) + \varepsilon_X(t+1) + \theta\varepsilon_X(t),$$

where $\pi_{YY} + \pi_{XX} = 0.85$ and $\pi_{YX}\pi_{XY} - \pi_{XX}\pi_{YY} = 0.105$. The parameters θ and $Var(\varepsilon_X(t)) = \sigma^2_{\varepsilon_X}$ are the solutions to the following system:

$$(1+\theta^2)\sigma^2_{\varepsilon_X} = 1.6125,$$

$$\theta\sigma^2_{\varepsilon_X} = -0.35.$$

The set of possible solutions is $\{(\theta,\, \sigma^2_{\varepsilon_X}) = (-4.378,\, 0.08),\ (-0.2285,\, 1.53)\}$. To get an invertible solution we must choose the combination which satisfies the condition $\mid \theta \mid < 1$, i.e. the combination $(-0.2285, 1.53)$. Thus, the variance of the forecast error of X at horizon 1 using only its own past is given by: $\Sigma[X(t+1) \mid \underline{X}_t] = 1.53$, and the variance of the forecast error of X at horizon 2 is $\Sigma[X(t+2) \mid \underline{X}_t] = 2.12$.

Table 1.1: Evaluation by simulation of causality at h=1, 2

p	$C(Y \underset{1}{\longrightarrow} X)$	$C(Y \underset{2}{\longrightarrow} X)$
1	0.519	0.567
2	0.430	0.220
3	0.427	0.200
4	0.425	0.199
5	0.426	0.198
10	0.425	0.197
15	0.426	0.199
20	0.425	0.197
25	0.425	0.199
30	0.426	0.198
35	0.425	0.198

Consequently, we have:

$$C(Y \underset{1}{\longrightarrow} X) = 0.425, \;\; C(Y \underset{2}{\longrightarrow} X) = 0.197.$$

In a second step we use the algorithm described at the beginning of this section to evaluate the causality measures using a large simulation technique. Table 1.1 shows results that we get for different lag orders p in the constrained model.[7] These results confirm the convergence ensured by the law of large numbers.

Now consider Example 2 of section 1:

$$\begin{bmatrix} X(t+1) \\ Y(t+1) \\ Z(t+1) \end{bmatrix} = \begin{bmatrix} 0.60 & 0.00 & 0.80 \\ 0.00 & 0.40 & 0.00 \\ 0.00 & 0.60 & 0.10 \end{bmatrix} \begin{bmatrix} X(t) \\ Y(t) \\ Z(t) \end{bmatrix} + \begin{bmatrix} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \\ \varepsilon_Z(t+1) \end{bmatrix} \tag{1.32}$$

In Example 2, analytical calculation of the causality measures at horizons 1 and 2 is not easy. In this example Y does not cause X at horizon 1, but causes it at horizon 2

[7]We consider $T = 600000$ simulations.

Table 1.2: Evaluation by simulation of causality at h=1, 2: Indirect causality

p	$C(Y \underset{1}{\longrightarrow} X \mid Z)$	$C(Y \underset{2}{\longrightarrow} X \mid Z)$
1	0.000	0.121
2	0.000	0.123
3	0.000	0.122
4	0.000	0.123
5	0.000	0.124
10	0.000	0.122
15	0.000	0.122
20	0.000	0.122
25	0.000	0.124
30	0.000	0.122
35	0.000	0.122

(indirect causality). Consequently, we expect that causality measure from Y to X will be equal to zero at horizon 1 and different from zero at horizon 2. Using a large simulation technique and by considering different lag orders p in the constrained model, we get the results in Table 1.2. These results show clearly the presence of an indirect causality from Y to X.

1.8 Confidence intervals

In this section, we assume that the process of interest $W \equiv \{W(s) = (X(s), Y(s), Z(s))' : s \leq t)\}$ follows a VAR(p) model[8]

$$W(t) = \sum_{j=1}^{p} \pi_j W(t-j) + u(t) \tag{1.33}$$

[8]If W follows a VAR(∞) model, then one can use Inoue and Kilian's (2002) approach to get results that are similar to those developed in this section.

or equivalently,

$$(I_3 - \sum_{j=1}^{p} \pi_j L^j)W(t) = u(t),$$

where the polynomial $\Pi(z) = I_3 - \sum_{j=1}^{p} \pi_j z^j$ satisfies $\det[\Pi(z)] \neq 0$, for $z \in \mathbb{C}$ with $|z| \leq 1$, and $\{u(t)\}_{t=0}^{\infty}$ is a sequence of *i.i.d.* random variables.[9]. For a realization $\{W(1), \ldots, W(T)\}$ of process W, estimates of $\Pi = (\pi_1, \ldots, \pi_p)$ and Σ_u are given by the following equations:

$$\hat{\Pi} = \hat{\Gamma}_1' \hat{\Gamma}^{-1}, \ \hat{\Sigma}_u = \sum_{t=p+1}^{T} \hat{u}(t)\hat{u}(t)'/(T-p), \tag{1.34}$$

where $\hat{\Gamma} = (T-p)^{-1}\sum_{t=p+1}^{T} w(t)w(t)'$, for $w(t) = (W(t)', \ldots, W(t-p+1)')'$, $\hat{\Gamma}_1 = (T-p)^{-1}\sum_{t=p+1}^{T} w(t)W(t+1)'$, and $\hat{u}(t) = W(t) - \sum_{j=1}^{p} \hat{\pi}_j W(t-j)$.

Now, suppose that we are interested in measuring causality from Y to X at given horizon h. In this case we need to know the structure of the marginal process $\{S(s) = (X(s), Z(s))', s \leq t)\}$. This one has a VARMA$(\bar{p}, \bar{q})$ representation with $\bar{p} \leq 3p$ and $\bar{q} \leq 2p$,

$$S(t) = \sum_{j=1}^{\bar{p}} \phi_j S(t-j) + \sum_{i=1}^{\bar{q}} \theta_i \varepsilon(t-i) + \varepsilon(t) \tag{1.35}$$

where $\{\varepsilon(t)\}_{t=0}^{\infty}$ is a sequence of i.i.d. random variables that satisfies

$$E[\varepsilon(t)] = 0, \ E\left[\varepsilon(t)\varepsilon'(s)\right] = \begin{cases} \Sigma_\varepsilon & \text{if } s = t \\ 0 & \text{if } s \neq t \end{cases},$$

and Σ_ε is a positive definite matrix. Equation (1.35) can be written in the following reduced form,

$$\phi(L)S(t) = \theta(L)\varepsilon(t),$$

where $\phi(L) = I_2 - \phi_1 L - \ldots - \phi_{\bar{p}} L^{\bar{p}}$ and $\theta(L) = I_2 + \theta_1 L + \ldots + \theta_{\bar{q}} L^{\bar{q}}$. We assume that

[9] We assume that X, Y, and Z are univariate variables. However, it is easy to generalize the results of this section to the multivariate case.

$\theta(z) = I_2 + \sum_{j=1}^{\bar{q}} \theta_j z^j$ satisfies $\det[\theta(z)] \neq 0$ for $z \in \mathbb{C}$ and $\mid z \mid \leq 1$. Under the latter assumption, the VARMA($\bar{p}, \bar{q}$) process is invertible and has a VAR(∞) representation:

$$S(t) - \sum_{j=1}^{\infty} \pi_j^c S(t-j) = \theta(L)^{-1}\phi(L)S(t) = \varepsilon(t). \tag{1.36}$$

Let $\Pi^c = (\pi_1^c, \pi_2^c, ..)$ denote the matrix of all autoregressive coefficients in model (1.36) and $\Pi^c(k) = (\pi_1^c, \pi_2^c, \ldots, \pi_k^c)$ denote its first k autoregressive coefficients. Suppose that we approximate (1.36) by a finite order VAR(k) model, where k depends on sample size T. The estimators of the autoregressive coefficients $\Pi^c(k)$ and variance-covariance matrix Σ_ε are given by:

$$\hat{\Pi}^c(k) = (\hat{\pi}_{1k}^c, \hat{\pi}_{2k}^c, \ldots, \hat{\pi}_{kk}^c) = \hat{\Gamma}_{k1}'\hat{\Gamma}_k^{-1}, \quad \hat{\Sigma}_{\varepsilon k} = \sum_{t=k+1}^{T} \hat{\varepsilon}_k(t)\hat{\varepsilon}_k(t)'/(T-k).$$

where $\hat{\Gamma}_k = (T-k)^{-1}\sum_{t=k+1}^{T} S_k(t)S_k(t)'$, for $S_k(t) = (S'(t), ..., S'(t-k+1))'$, $\hat{\Gamma}_{k1} = (T-k)^{-1}\sum_{t=k+1}^{T} S_k(t)S(t+1)'$, and $\hat{\varepsilon}_k(t) = S(t) - \sum_{j=1}^{k} \hat{\pi}_{jk}^c S(t-j)$.

From the above notations, the theoretical value of the causality measure from Y to X at horizon h may be defined as follows:

$$C(Y \underset{h}{\longrightarrow} X|Z) = \ln\left(\frac{G(\text{vec}(\Pi^c), \text{vech}(\Sigma_\varepsilon))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))}\right)$$

where

$$G(\text{vec}(\Pi^c), \text{vech}(\Sigma_\varepsilon)) = \sum_{j=0}^{h-1} e_c' \pi_1^{c(j)} \Sigma_\varepsilon \pi_1^{c(j)'} e_c, \quad e_c = (1, 1)',$$

$$H(\text{vec}(\Pi), \text{vech}(\Sigma_u)) = \sum_{j=0}^{h-1} e_{nc}' \pi_1^{(j)} \Sigma_u \pi_1^{(j)'} e_{nc}, \quad e_{nc} = (1, 1, 1)',$$

with $\pi_1^{c(j)} = \pi_2^{c(j-1)} + \pi_1^{c(j-1)}\pi_1^c$, for $j \geq 2$, $\pi_1^{c(0)} = I_2$, and $\pi_1^{c(1)} = \pi_1^c$ [see Dufour and Renault (1998)]. vec denotes the column stacking operator and vech is the column stacking operator that stacks the elements on and below the diagonal only. By Corollary 2, there exists a function $f(.) : \mathbb{R}^{9(p+1)} \longrightarrow \mathbb{R}^{4(k+1)}$ which associates the constrained parameters $(\text{vec}(\Pi^c), \text{vech}(\Sigma_\varepsilon))$ with the unconstrained parameters $(\text{vec}(\Pi), \text{vech}(\Sigma_u))$ such that [see Example 4]:

$$(\text{vec}(\Pi^c), \text{vech}(\Sigma_\varepsilon))' = f((\text{vec}(\Pi), \text{vech}(\Sigma_u))')$$

and

$$C(Y \underset{h}{\longrightarrow} X|Z) = \ln\left(\frac{G(f((\text{vec}(\Pi), \text{vech}(\Sigma_u))'))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))}\right).$$

An estimator of $C(Y \underset{h}{\longrightarrow} X \mid Z)$ is given by:

$$\hat{C}(Y \underset{h}{\longrightarrow} X|Z) = \ln\left(\frac{G(\text{vec}(\hat{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_{\varepsilon,k}))}{H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u))}\right). \tag{1.37}$$

where $G(\text{vec}(\hat{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_{\varepsilon,k}))$ and $H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u))$ are estimates of the corresponding population quantities.

Now let consider the following assumptions.

Assumption 1 : [see Lewis and Reinsel (1985)]

1) $E \mid \varepsilon_h(t)\varepsilon_i(t)\varepsilon_j(t)\varepsilon_l(t) \mid \leq \gamma_4 < \infty$, for $1 \leq h, i, j, l \leq 2$;

2) k is chosen as a function of T such that $k^3/T \to 0$ as $k, T \to \infty$;

3) k is chosen as a function of T such that $T^{1/2} \sum_{j=k+1}^{\infty} \| \pi_j^c \| \to 0$ as $k, T \to \infty$.

4) Series used to estimate parameters of $VAR(k)$ and series used for prediction are generated from two independent processes which have the same stochastic structure.

Assumption 2 : $f(.)$ is continuous and differentiable *function.*

Proposition 6 (Consistency of $\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)$) *Under assumption (1), $\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)$ is a consistent estimator of $C(Y \underset{h}{\longrightarrow} X \mid Z)$.*

To establish the asymptotic distribution of $\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)$, let us start by recalling the following result [see Lütkepohl (1990a, page 118-119) and Kilian (1998a, page 221)]:

$$T^{1/2}\begin{pmatrix} \text{vec}(\hat{\Pi}) - \text{vec}(\Pi) \\ \text{vech}(\hat{\Sigma}_u) - \text{vech}(\Sigma_u) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Omega) \tag{1.38}$$

where

$$\Omega = \begin{pmatrix} \Gamma^{-1} \otimes \Sigma_u & 0 \\ 0 & 2(D_3'D_3)^{-1}D_3'(\Sigma_u \otimes \Sigma_u)D_3(D_3'D_3)^{-1} \end{pmatrix}.$$

D_3 is the duplication matrix, defined such that $\text{vech}(F) = D_3\text{vech}(F)$ for any symmetric 3×3 matrix F.

Proposition 7 (Asymptotic distribution of $\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)$) *Under assumptions (1) and (2), we have:*

$$T^{1/2}[\hat{C}(Y \underset{h}{\longrightarrow} X|Z) - C(Y \underset{h}{\longrightarrow} X|Z)] \xrightarrow{d} \mathcal{N}(0, \Sigma_C)$$

where $\Sigma_C = D_C \Omega D_C'$, *and*

$$D_C = \frac{\partial C(Y \underset{h}{\longrightarrow} X|Z)}{\partial(\text{vec}(\Pi)', \text{vech}(\Sigma_u)')}.$$

$$\Omega = \begin{pmatrix} \Gamma^{-1} \otimes \Sigma_u & 0 \\ 0 & 2(D_3'D_3)^{-1}D_3'(\Sigma_u \otimes \Sigma_u)D_3(D_3'D_3)^{-1} \end{pmatrix}.$$

Analytically differentiating the causality measure with respect to the vector $(\text{vec}(\Pi), \text{vech}(\Sigma_u))'$ is not feasible. One way to build confidence intervals for causality measures is to use a large simulation technique [see section 2.4] to calculate the derivative numerically. Another way is by building bootstrap confidence intervals. As mentioned by Inoue and Kilian (2002), for bounded measures, as in our case, the bootstrap approach is more reliable than the delta-method. The reason is because the delta-method interval is not range respecting and may produce confidence intervals that are logically invalid. In contrast, the bootstrap percentile interval by construction preserves these constraints [see Inoue and Kilian (2002) and Efron and Tibshirani (1993)].

Let us consider the following bootstrap approximation to the distribution of the causality measure at given horizon h.

1. Estimate a VAR(p) process and save the residuals

$$\tilde{u}(t) = W(t) - \sum_{j=1}^{p} \hat{\pi}_j W(t-j), \text{ for } t = p+1, \ldots, T,$$

   where $\hat{\pi}_j$, for $j = 1, \ldots, p$, are given by equation (1.34).

2. Generate $(T-p)$ bootstrap residuals $\tilde{u}^*(t)$ by random sampling with replacement from the residuals $\tilde{u}(t)$, $t = p+1, \ldots, T$.

3. Choose the vector of p initial observations $w(0) = (W'(1), \ldots, W'(p))'$.[10]

4. Given $\hat{\Pi}=(\hat{\pi}_1, \ldots, \hat{\pi}_p)$, $\tilde{u}^*(t)$, and $w(0)$, generate bootstrap data for the dependent variable $W^*(t)$ from equation:

$$W^*(t) = \sum_{j=1}^{p} \hat{\pi}_j W^*(t-j) + \tilde{u}^*(t), \text{ for } t = p+1, \ldots, T. \tag{1.39}$$

5. Calculate the bootstrap OLS regression estimates

$$\hat{\Pi}^* = (\hat{\pi}_1^*, \hat{\pi}_2^*, \ldots, \hat{\pi}_p^*) = \hat{\Gamma}_1^{*\prime}\hat{\Gamma}^{*-1}, \quad \hat{\Sigma}_u^* = \sum_{t=p+1}^{T} \tilde{u}^*(t)\tilde{u}^*(t)'/(T-p),$$

   where $\hat{\Gamma}^* = (T-p)^{-1}\sum_{t=p+1}^{T} w^*(t)w^*(t)'$, for $w^*(t) = (W^{*\prime}(t), \ldots, W^{*\prime}(t-p+1))'$, $\hat{\Gamma}_1^* = (T-p)^{-1}\sum_{t=p+1}^{T} w^*(t)W^*(t+1)'$, and $\tilde{u}^*(t) = W^*(t) - \sum_{j=1}^{p} \hat{\pi}_j W^*(t-j)$.

[10] The choice of using the initial vectors $(W'(1), \ldots, W'(p))'$ seems natural, but any block of p vectors from $W \equiv \{W(1), \ldots, W(T)\}$ would be appropriate. Berkowitz and Kilian (2000) note that conditioning each bootstrap replicate on the same initial value will understate the uncertainty associated with the bootstrap estimates, and this choice is randomised in the simulations by choosing the starting value from $W \equiv \{W(1), \ldots, W(T)\}$ [see Patterson (2007)].

6. Estimate the constrained model of the marginal process (X, Z) using the bootstrap sample $\{W^*(t)\}_{t=1}^T$.

7. Calculate the causality measure at horizon h, denoted $\hat{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z)$, using equation (1.37).

8. Choose B such $\frac{1}{2}\alpha(B+1)$ is an integer and repeat steps $(2)-(7)$ B times.

Conditional on the sample, we have [see Inoue and Kilian (2002)],

$$T^{1/2}\begin{pmatrix} \text{vec}(\hat{\Pi}^*) - \text{vec}(\hat{\Pi}) \\ \text{vech}(\hat{\Sigma}_u^*) - \text{vech}(\hat{\Sigma}_u) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Omega), \tag{1.40}$$

where

$$\Omega = \begin{pmatrix} \Gamma^{-1} \otimes \Sigma_u & 0 \\ 0 & 2(D_3' D_3)^{-1} D_3' (\Sigma_u \otimes \Sigma_u) D_3 (D_3' D_3)^{-1} \end{pmatrix}.$$

D_3 is the duplication matrix defined such that $\text{vech}(F) = D_3 \text{vech}(F)$ for any symmetric 3×3 matrix F. We have the following result which establish the validity of the percentile bootstrap technique.

Proposition 8 (Asymptotic validity of the residual-based bootstrap) *Under assumptions (1) and (2), we have*

$$T^{1/2}(\hat{C}^*(Y \underset{h}{\longrightarrow} X|Z) - \hat{C}(Y \underset{h}{\longrightarrow} X|Z)) \xrightarrow{d} \mathcal{N}(0, \Sigma_C).$$

where $\Sigma_C = D_C \Omega D_C'$ *and*

$$D_C = \frac{\partial C(Y \underset{h}{\longrightarrow} X|Z)}{\partial(\text{vec}(\Pi)', \text{vech}(\Sigma_u)')}.$$

$$\Omega = \begin{pmatrix} \Gamma^{-1} \otimes \Sigma_u & 0 \\ 0 & 2(D_3' D_3)^{-1} D_3' (\Sigma_u \otimes \Sigma_u) D_3 (D_3' D_3)^{-1} \end{pmatrix}.$$

Kilian (1998) proposes an algorithm to remove the bias in impulse response functions prior to bootstrapping the estimate. As he mentioned, the small sample bias in an impulse response function may arise from bias in slope coefficient estimates or from the nonlinearity of this function, and this can translate into changes in interval width and location. If the ordinary least-squares small-sample bias can be responsible for bias in the estimated impulse response function, then replacing the biased slope coefficient estimates by bias-corrected slope coefficient estimates may help to reduce the bias in the impulse response function. Kilian (1998) shows that the additional modifications proposed in the bias-corrected bootstrap confidence intervals method do not alter its asymptotic validity. The reason is that the effect of bias corrections is negligible asymptotically.

To improve the performance of the percentile bootstrap intervals described above, we almost consider the same algorithm as in Kilian (1998). Before bootstrapping the causality measures, we correct the bias in the VAR coefficients. We approximate the bias term $Bias = E[\hat{\Pi} - \Pi]$ of the VAR coefficients by the corresponding bootstrap bias $Bias^* = E^*[\hat{\Pi}^* - \hat{\Pi}]$, where E^* is the expectation based on the bootstrap distribution of $\hat{\Pi}^*$. This suggests the bias estimate

$$\widehat{Bias}^* = \frac{1}{B}\sum_{j=1}^{B} \hat{\Pi}^{*(j)} - \hat{\Pi}.$$

We substitute $\hat{\Pi} - \widehat{Bias}^*$ in equation (1.39) and generate B new bootstrap replications $\hat{\Pi}^*$. We use the same bias estimate, $\widehat{Bias}^*$, to estimate the mean bias of new $\hat{\Pi}^*$.[11] Then we calculate the bias-corrected bootstrap estimator $\tilde{\Pi}^* = \hat{\Pi}^* - \widehat{Bias}^*$ that we use to estimate the bias-corrected bootstrap causality measure estimate. Based on the discussion by Kilian (1998, page 219), given the nonlinearity of the causality measure, this procedure will not in general produce unbiased estimates, but as long as the resulting bootstrap estimator is approximately unbiased, the implied percentile intervals are likely to be good approximations. To reduce more the bias in the causality measures estimate,

[11] See Kilian (1998).

in our empirical application we consider another bias correction directly on the measure itself, this one is given by

$$\tilde{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z) = \hat{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z) - [\overline{C}^{*}(Y \underset{h}{\longrightarrow} X \mid Z) - \hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)],$$

where

$$\overline{C}^{*}(Y \underset{h}{\longrightarrow} X \mid Z) = \frac{1}{B}\sum_{j=1}^{B}\tilde{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z).$$

In practice, specially when the true value of causality measure is close to zero, it is possible that for some bootstrap samples

$$\hat{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z) \leq [\overline{C}^{*}(Y \underset{h}{\longrightarrow} X \mid Z) - \hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)],$$

in this case we impose the following non-negativity truncation:

$$\tilde{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z) = \max\left\{\tilde{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z), 0\right\}.$$

1.9 Empirical illustration

In this section, we apply our causality measures to measure the strength of relationships between macroeconomic and financial variables. The data set considered is the one used by Bernanke and Mihov (1998) and Dufour, Pelletier, and Renault (2006). This data set consists of monthly observations on nonborrowed reserves (NBR), the federal funds rate (r), the gross domestic product deflator (P), and real gross domestic product (GDP). The monthly data on GDP and the GDP deflator were constructed using state space methods from quarterly observations [for more details, see Bernanke and Mihov (1998)]. The sample runs from January 1965 to December 1996 for a total of 384 observations.

All variables are in logarithmic form [see Figures 1–4]. These variables were also transformed by taking first differences [see Figures 5–8], consequently the causality relations have to be interpreted in terms of the growth rate of variables.

Figure 1: NBR in logarithmic form

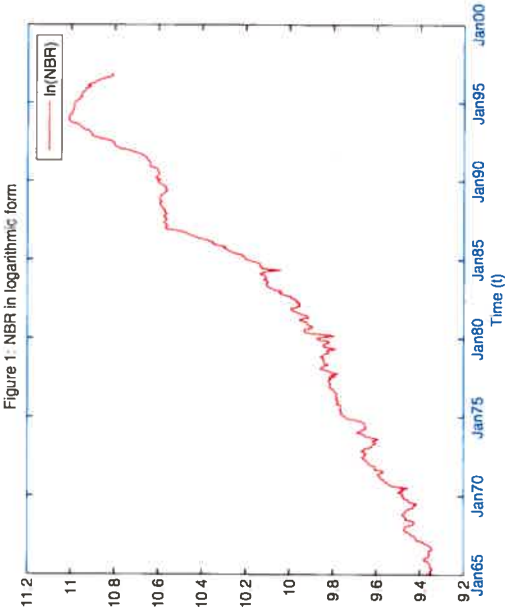


Figure 2: r in logarithmic form

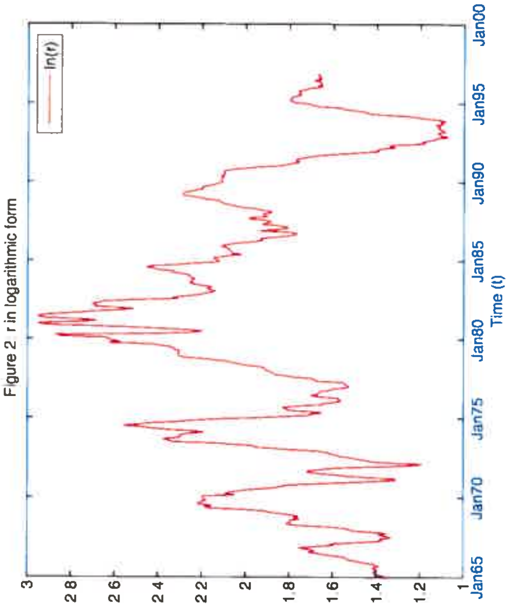


Figure 3. P in logarithmic form

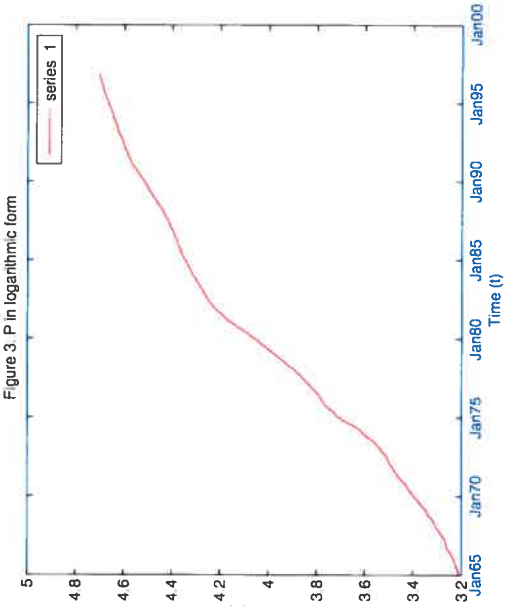


Figure 4: GDP in logarithmic form

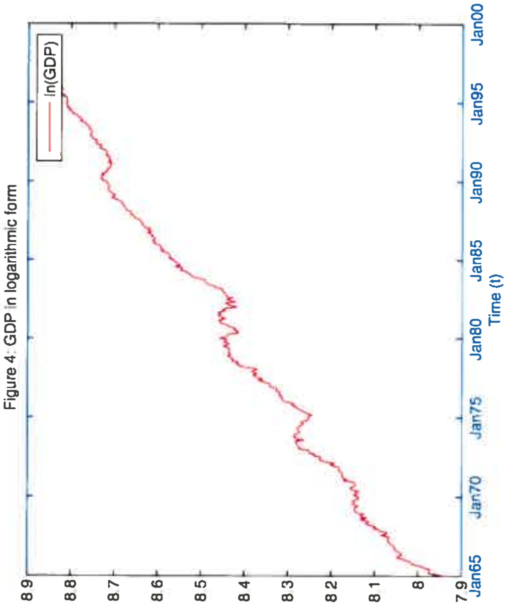


Figure 5 The first differentiation of ln(NBR)

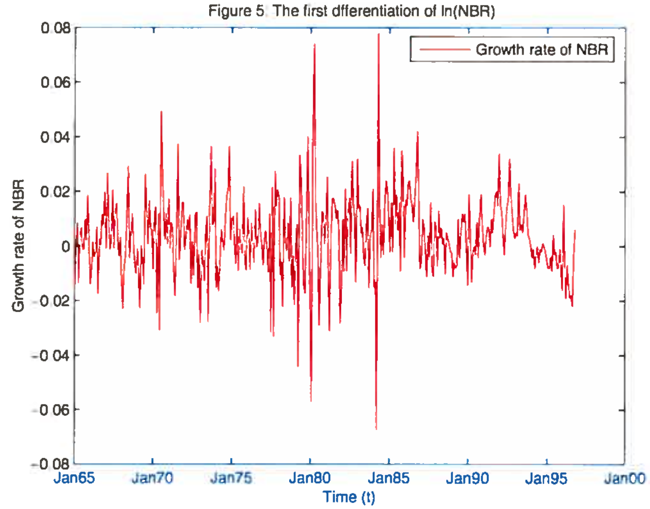


Figure 6 The first differentiation of ln(r)

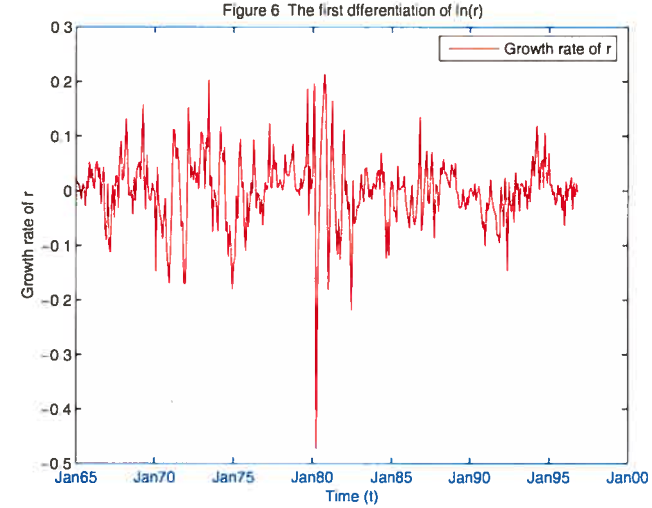


Figure 7 The first differentiation of ln(P)

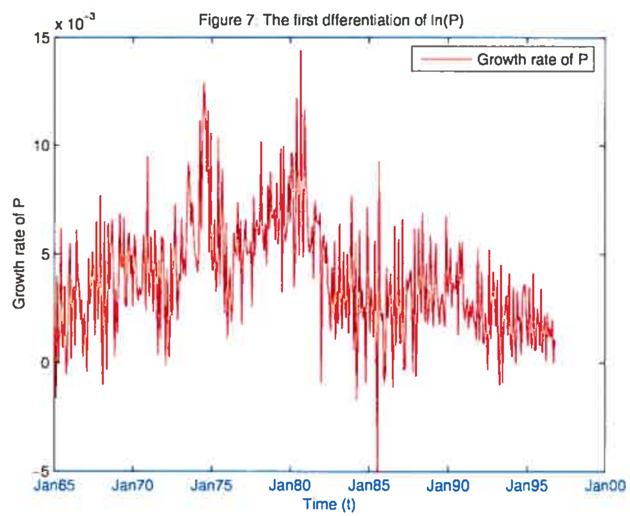


Figure 8 The first differentiation of ln(GDP)

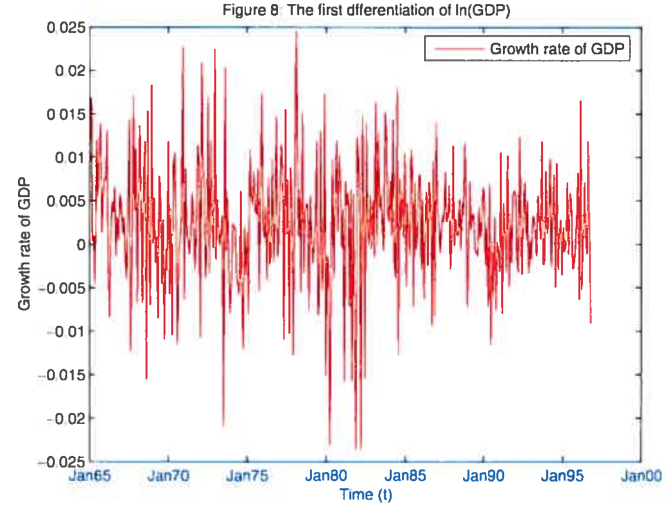


Table 1.3: Dickey-Fuller tests: Variables in Logarithmic form

	With Intercept		With Intercept and Trend	
	ADF test statistic	5% Critical Value	ADF test statistic	5% Critical Value
NBR	−0.510587	−2.8694	−1.916428	−3.4234
R	−2.386082	−2.8694	−2.393276	−3.4234
P	−1.829982	−2.8694	−0.071649	−3.4234
GDP	−1.142940	−2.8694	−3.409215	−3.4234

Table 1.4: Dickey-Fuller tests: First difference

	With Intercept		With Intercept and Trend	
	ADF test statistic	5% Critical Value	ADF test statistic	5% Critical Value
NBR	−5.956394	−2.8694	−5.937564	−3.9864
r	−7.782581	−2.8694	−7.817214	−3.9864
P	−2.690660	−2.8694	−3.217793	−3.9864
GDP	−5.922453	−2.8694	−5.966043	−3.9864

We performed an Augmented Dickey-Fuller test (hereafter ADF-test) for nonstationarity of the four variables of interest and their first differences. The values of the test statistics, as well as the critical values corresponding to a 5% significance level, are given in tables 1.3 and 1.4. Table 1.5, below, summarizes the results of the stationarity tests for all variables.

As we can read from Table 1.5, all variables in logarithmic form are nonstationary. How-

Table 1.5: Stationarity test results

	Variables in logarithmic form	First difference
NBR	No	Yes
r	No	Yes
P	No	No
GDP	No	Yes

ever, their first differences are stationary except for the GDP deflator, P. We performed a nonstationarity test for the second difference of variable P. The test statistic values are equal to -11.04826 and -11.07160 for the ADF-test with only an intercept and with both intercept and trend, respectively. The critical values in both cases are equal to -2.8695 and -3.4235. Thus, the second difference of variable P is stationary.

Once the data is made stationary, we use a nonparametric approach for the estimation and Akaike's information criterion to specify the orders of the long VAR(k) models. To choose the upper bound on the admissible lag orders K, we apply the results in Lewis and Reinsel (1985). Using Akaike's criterion for the unconstrained VAR model, which corresponds to four variables, we observe that it is minimized at $k = 16$. We use same criterion to specify the orders of the constrained VAR models, which correspond to different combinations of three variables, and we find that the orders are all less than or equal to 16. To compare the determinants of the variance-covariance matrices of the constrained and unconstrained forecast errors at horizon h, we take the same order $k = 16$ for the constrained and unconstrained models. We compute different causality measures for horizons $h = 1, \ldots, 40$ [see Figures 9–14]. Higher values of measures indicate greater causality. We also calculate the corresponding nominal 95% bootstrap confidence intervals as described in the previous section.

From Figure 9 we observe that nonborrowed reserves have a considerable effect on the federal funds rate at horizon one comparatively with other variables [see Figures 10 and 11]. This effect is well known in the literature and can be explained by the theory of supply and demand for money. We also note that nonborrowed reserves have a short-term effect on GDP and can cause the GDP deflator until horizon 5. Figure 14 shows the effect of GDP on the federal funds rate is significant for the first four horizons. The effect of the federal funds rate on the GDP deflator is significant only at horizon 1 [see Figure 12]. Other significant results concern the causality from r to GDP. Figure 13 shows the effect of the interest rate on GDP is significant until horizon 16. These results are consistent with conclusions obtained by Dufour et al. (2005).

Table 6 represents results of other causality directions until horizon 20. As we can read from this table, there is no causality in these other directions. Finally, note that the above results do not change when we consider the second, rather than first, difference of variable P.

1.10 Conclusion

New concepts of causality were introduced in Dufour and Renault (1998): causality at a given (arbitrary) horizon h, and causality up to any given horizon h, where h is a positive integer and can be infinite ($1 \leq h \leq \infty$). These concepts are motivated by the fact that, in the presence of an auxiliary variable Z, it is possible to have a situation in which the variable Y does not cause variable X at horizon 1, but causes it at a longer horizon $h > 1$. In this case, this is an indirect causality transmitted by the auxiliary variable Z.

Another related problem arises when measuring the importance of the causality between two variables. Existing causality measures have been established only for horizon 1 and fail to capture indirect causal effects. This chapter proposes a generalization of such measures for any horizon h. We propose parametric and nonparametric measures for feedback and instantaneous effects at any horizon h. Parametric measures are defined in terms of impulse response coefficients in the VMA representation. By analogy with Geweke (1982), we show that it is possible to define a measure of dependence at horizon h which can be decomposed into a sum of feedback measures from X to Y, from Y to X, and an instantaneous effect at horizon h. We also show how these causality measures can be related to the predictability measures developed in Diebold and Kilian (1998).

We propose a new approach to estimating these measures based on simulating a large sample from the process of interest. We also propose a valid nonparametric confidence interval, using the bootstrap technique.

From an empirical application we found that nonborrowed reserves cause the federal funds rate only in the short term, the effect of real gross domestic product on the federal funds rate is significant for the first four horizons, the effect of the federal funds rate on the gross domestic product deflator is significant only at horizon 1, and finally the federal funds rate causes the real gross domestic product until horizon 16.

Figure 9: Causality measures from Nonborrowed reserves to Federal funds rate

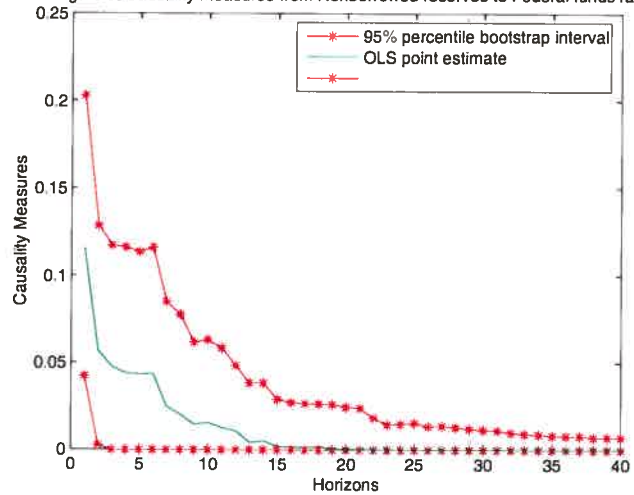


Figure10: Causality measures from Nonborrowed reserves to GDP Deflator

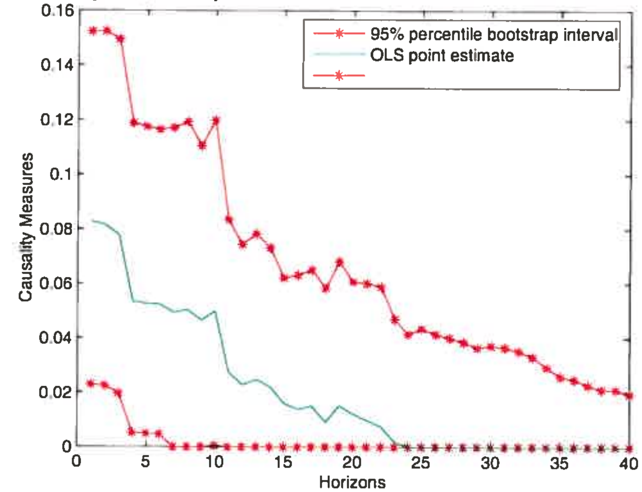


Figure 11: Causality measures from Nonborrowed reserves to Real GDP

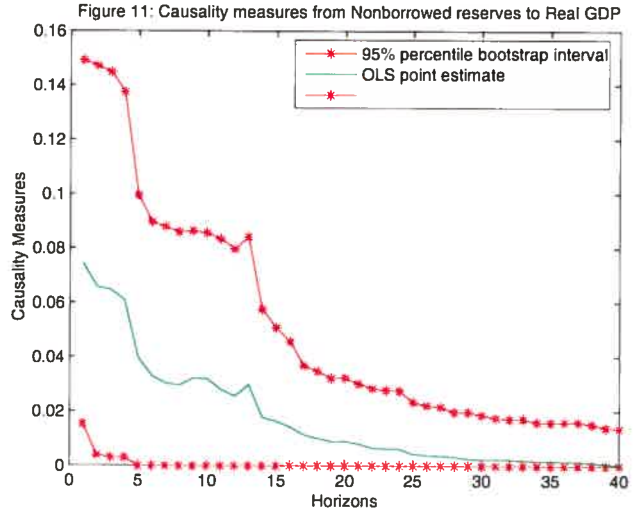


Figure 12: Causality measures from Federal funds rate to GDP Deflator

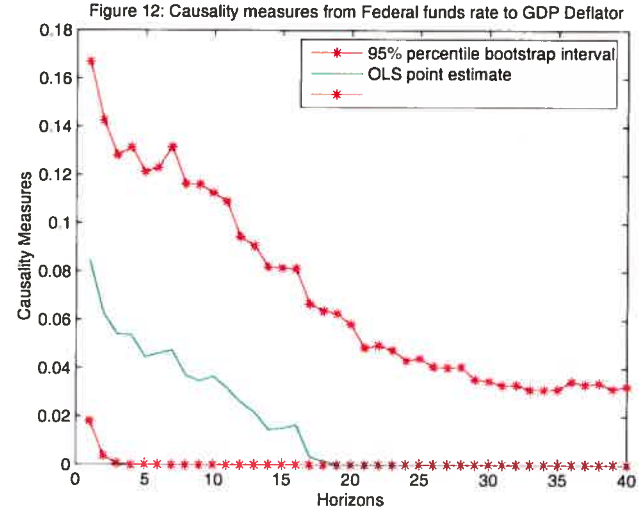


Figure 13: Causality measures from Federal funds rate to Real GDP

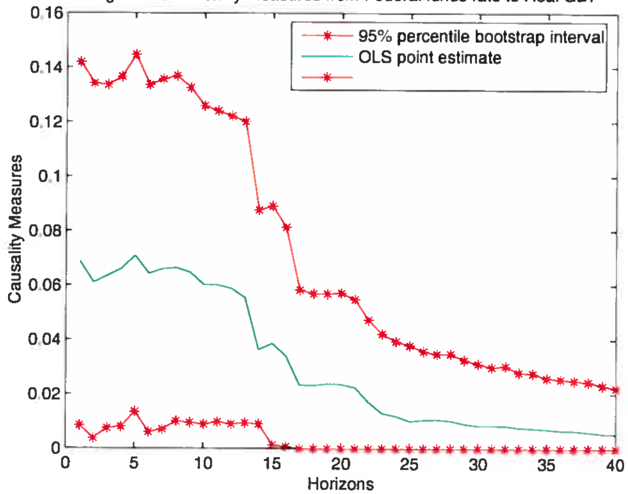


Figure 14: Causality measures from Real GDP to Federal funds rate

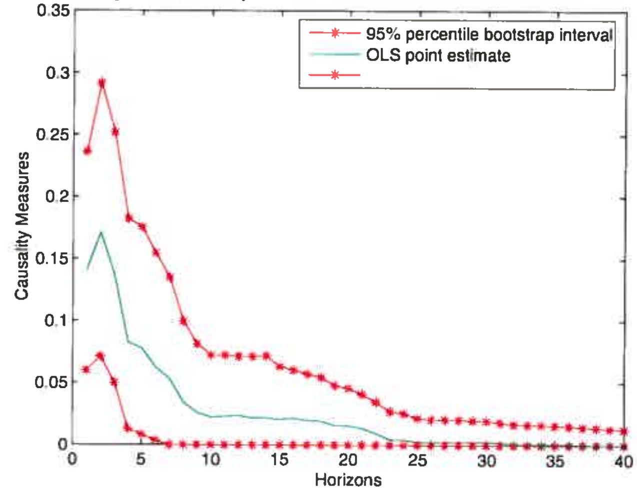


Table 6. Summary of causality relations at various horizons for series in first difference

h	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$NBR \to R$	yes																			
$NBR \to P$	yes	yes	yes	yes																
$NBR \to GDP$	yes																			
$R \to NBR$																				
$R \to P$	yes																			
$R \to GDP$	yes		yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes				
$P \to NBR$																				
$P \to R$																				
$P \to GDP$																				
$GDP \to NBR$																				
$GDP \to R$	yes	yes	yes	yes	yes															
$GDP \to P$																				

1.11 Appendix: Proofs

Proof of Proposition 1. From definition 3 and for $m_1 = m_2 = 1$,

$$
\begin{aligned}
C(Y \underset{h_2}{\longrightarrow} X|Z) &= \ln\left[\frac{\sigma^2(X(t+h_1) \mid I_t - \mathrm{Y}_t)}{\sigma^2(X(t+h_1) \mid I_t)}\right] \\
&+ \ln\left[\frac{\sigma^2(X(t+h_1) \mid I_t)\sigma^2(X(t+h_2) \mid I_t - \mathrm{Y}_t)}{\sigma^2(X(t+h_1) \mid I_t - \mathrm{Y}_t)\sigma^2(X(t+h_2) \mid I_t)}\right] \\
&= C(Y \underset{h_1}{\longrightarrow} X|Z) + \ln\left[\frac{\sigma^2(X(t+h_1) \mid I_t)}{\sigma^2(X(t+h_2) \mid I_t)}\right] - \ln\left[\frac{\sigma^2(X(t+h_1) \mid I_t - \mathrm{Y}_t)}{\sigma^2(X(t+h_2) \mid I_t - \mathrm{Y}_t)}\right].
\end{aligned}
$$

According to Diebold and Kilian (1998), the predictability measure of vector X under the information sets $I_t - \mathrm{Y}_t$ and I_t are, respectively, defined as follows:

$$
\begin{aligned}
\bar{P}_X(I_t - \mathrm{Y}_t, h_1, h_2) &= 1 - \frac{\sigma^2(X(t+h_1) \mid I_t - \mathrm{Y}_t)}{\sigma^2(X(t+h_2) \mid I_t - \mathrm{Y}_t)}, \\
\bar{P}_X(I_t, h_1, h_2) &= 1 - \frac{\sigma^2(X(t+h_1) \mid I_t - \mathrm{Y}_t)}{\sigma^2(X(t+h_2) \mid I_t - \mathrm{Y}_t)}.
\end{aligned}
$$

Hence the result to be proved:

$$
C(Y \underset{h_1}{\longrightarrow} X|Z) - C(Y \underset{h_2}{\longrightarrow} X|Z) = \ln\left[1 - \bar{P}_X(I_t - \mathrm{Y}_t, h_1, h_2)\right] - \ln\left[1 - \bar{P}_X(I_t, h_1, h_2)\right].
$$

■

Proof of Proposition 6. Under assumption 1 and using Theorem 6 in Lewis and Reinsel (1985),

$$
\begin{aligned}
G\big(\mathrm{vec}(\hat{\Pi}^c(k)), \mathrm{vech}(\hat{\Sigma}_{\varepsilon,k})\big) &= (1 + \frac{2k}{T}) G(\mathrm{vec}(\Pi^c), \mathrm{vech}(\Sigma_\varepsilon)) \\
&= (1 + O(T^{-\delta})) G(\mathrm{vec}(\Pi^c), \mathrm{vech}(\Sigma_\varepsilon)), \text{ for } \frac{2}{3} < \delta < 1.
\end{aligned}
$$

The second equality follows from condition 2 of assumption 1. If we consider that $k = T^\alpha$ for $\alpha > 0$, then condition 2 implies that $k^3/T = T^{3\alpha - 1}$ with $0 < \alpha < \frac{1}{3}$. Similarly, we have $\frac{2k}{T} = 2T^{\alpha-1}$ and $T^{\delta}(2T^{\alpha-1}) \rightarrow 2$ for $\frac{2}{3} < \delta < 1$. Thus, for $\frac{2}{3} < \delta < 1$,

$$\begin{aligned}\ln\left(G(\text{vec}(\hat{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_{\varepsilon, k}))\right) &= \ln\left(G(\text{vec}(\Pi^c), \text{vech}(\Sigma_{\varepsilon}))+\ln\left(1+O(T^{-\delta}\right)\right) \\ &= \ln\left(G(\text{vec}(\Pi^c), \text{vech}(\Sigma_{\varepsilon}))+O(T^{-\delta}\right). \qquad (1.41)\end{aligned}$$

For $H(.)$ a continuous function in $(\text{vec}(\Pi), \text{vech}(\Sigma_u))$ and because $\hat{\Pi} \underset{p}{\rightarrow} \Pi$, $\hat{\Sigma}_u \underset{p}{\rightarrow} \Sigma_u$, we have

$$\ln\left(H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u))\right) \underset{p}{\rightarrow} \ln\left(H(\text{vec}(\Pi), \text{vech}(\Sigma_u))\right). \qquad (1.42)$$

Thus, from (1.41)-(1.42) and for $\frac{2}{3} < \delta < 1$, we get

$$\hat{C}\,(Y \underset{h}{\longrightarrow} X \mid Z) = \ln\left(\frac{G(\text{vec}(\Pi^c), \text{vech}(\Sigma_{\varepsilon}))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))}\right) + O(T^{-\delta}) + o_p(1).$$

Consequently,

$$\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z) \underset{p}{\rightarrow} C(Y \underset{h}{\longrightarrow} X \mid Z).$$

■

Proof of Proposition 7. We have shown [see proof of consistency] that, for $\frac{2}{3} < \delta < 1$,

$$\ln\left(G(\text{vec}(\hat{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_{\varepsilon, k}))\right) = \ln\left(G(f(\text{vec}(\Pi), \text{vech}(\Sigma_u)))\right) + O(T^{-\delta}). \qquad (1.43)$$

Under Assumption 2, we have

$$\ln(G(f(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))) \underset{p}{\rightarrow} \ln(G(f(\text{vec}(\Pi), \text{vech}(\Sigma_u)))). \qquad (1.44)$$

Thus, from (1.43)–(1.44) and for $\frac{2}{3} < \delta < 1$, we get:

$$\ln\left(G(\text{vec}(\hat{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_{\varepsilon, k}))\right) = \ln\left(G(f(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))\right) + O(T^{-\delta}) + o_p(1).$$

Consequently, for $\frac{2}{3} < \delta < 1$,

$$\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z) = \ln\left(\frac{G(f(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))}{H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u))}\right) + O(T^{-\delta}) + o_p(1)$$

$$= \tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) + O(T^{-\delta}) + o_p(1)$$

where

$$\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) = \ln\left(\frac{G(f(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))}{H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u))}\right).$$

Since

$$\ln\left(\frac{G(f(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))}{H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u))}\right) = O_p(1),$$

the asymptotic distribution of $\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)$ will be the same as that of $\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z)$. Furthermore, using Assumption 2 and a first-order Taylor expansion of $\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z)$, we have:

$$\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) = C(Y \underset{h}{\longrightarrow} X \mid Z) + D_C \begin{pmatrix} \text{vec}(\hat{\Pi}) - \text{vec}(\Pi) \\ \text{vech}(\hat{\Sigma}_u) - \text{vech}(\Sigma_u) \end{pmatrix} + o_p(T^{-\frac{1}{2}}),$$

where

$$D_C = \frac{\partial C(Y \underset{h}{\longrightarrow} X \mid Z)}{\partial(\text{vec}(\Pi)', \text{vech}(\Sigma_u)')},$$

hence

$$T^{1/2}[\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) - C(Y \underset{h}{\longrightarrow} X \mid Z)] \simeq D_C \begin{pmatrix} T^{1/2}\,\text{vec}(\hat{\Pi}) - \text{vec}(\Pi) \\ T^{1/2}\text{vech}(\hat{\Sigma}_u) - \text{vech}(\Sigma_u) \end{pmatrix}.$$

From (1.38), we have

$$T^{1/2}[\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) - C(Y \underset{h}{\longrightarrow} X \mid Z)] \xrightarrow{d} \mathcal{N}(0, \Sigma_C).$$

hence

$$T^{1/2\prime}[\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z) - C(Y \underset{h}{\longrightarrow} X|Z)] \xrightarrow{d} \mathcal{N}(0, \Sigma_C)$$

where

$$\Sigma_C = D_C \Omega D_C',$$

$$\Omega = \begin{pmatrix} \Gamma^{-1} \otimes \Sigma_u & 0 \\ 0 & 2(D_3' D_3)^{-1} D_3' (\Sigma_u \otimes \Sigma_u) D_3 (D_3' D_3)^{-1} \end{pmatrix}.$$

D_3 is the duplication matrix, defined such that $\text{vech}(F) = D_3 \text{vech}(F)$ for any symmetric 3×3 matrix F. ■

Proof of Proposition 8. We start by showing that

$$\text{vec}(\hat{\Pi}^*) \underset{p}{\rightarrow} \text{vec}(\hat{\Pi}), \ \text{vech}(\hat{\Sigma}_u^*) \underset{p}{\rightarrow} \text{vech}(\hat{\Sigma}_u),$$
$$\text{vec}(\hat{\Pi}^{c*}(k)) \underset{p}{\rightarrow} \text{vec}(\hat{\Pi}^c(k)), \ \text{vech}(\hat{\Sigma}_{\varepsilon,k}^*) \underset{p}{\rightarrow} \text{vech}(\hat{\Sigma}_{\varepsilon,k}).$$

We first note that

$$\begin{aligned}
\text{vec}(\hat{\Pi}^*) &= \text{vec}(\hat{\Gamma}_1^{*\prime} \hat{\Gamma}^{*-1}) = \text{vec}((T-p)^{-1} \sum_{t=p+1}^{T} W(t+1)^* w^*(t)' \hat{\Gamma}^{*-1}) \\
&= \text{vec}((T-p)^{-1} \sum_{t=p+1}^{T} [\hat{\Pi} w^*(t) + \tilde{u}^*(t+1)] w^*(t)' \hat{\Gamma}^{*-1}) \\
&= \text{vec}(\hat{\Pi}((T-p)^{-1} \sum_{t=p+1}^{T} w^*(t) w^*(t)') \hat{\Gamma}^{*-1}) \\
&+ \text{vec}((T-p)^{-1} \sum_{t=p+1}^{T} \tilde{u}^*(t+1) w^*(t)' \hat{\Gamma}^{*-1}) \\
&= \text{vec}(\hat{\Pi}\ \hat{\Gamma}^* \hat{\Gamma}^{*-1}) + \text{vec}((T-p)^{-1} \sum_{t=p+1}^{T} \tilde{u}^*(t+1) w^*(t)' \hat{\Gamma}^{*-1}).
\end{aligned}$$

Let $\Im_t^* = \sigma(\tilde{u}^*(1), \ldots, \tilde{u}^*(t))$ denote the σ-algebra generated by $\tilde{u}^*(1), \ldots, \tilde{u}^*(t)$. Then,

$$\begin{aligned} \mathsf{E}^*[\tilde{u}^*(t+1)w^*(t)'\hat{\Gamma}^{*-1}] &= \mathsf{E}^*[\mathsf{E}^*[\tilde{u}^*(t+1)w^*(t)'\hat{\Gamma}^{*-1} \mid \Im_t^*]] \\ &= \mathsf{E}^*[\mathsf{E}^*[\tilde{u}^*(t+1) \mid \Im_t^*]w^*(t)'\hat{\Gamma}^{*-1}] = 0. \end{aligned}$$

By the law of large numbers,

$$(T-p)^{-1} \sum_{t=p+1}^{T} \tilde{u}^*(t+1)w^*(t)'\hat{\Gamma}^{*-1} = \mathsf{E}^*[\tilde{u}^*(t+1)w^*(t)'\hat{\Gamma}^{*-1}] + o_p(1)\,.$$

Thus,

$$\mathrm{vec}(\hat{\Pi}^*) - \mathrm{vec}(\hat{\Pi}) \underset{p}{\rightarrow} 0\,.$$

Now, to prove that $\mathrm{vech}(\hat{\Sigma}_u^*) \underset{p}{\rightarrow} \mathrm{vech}(\hat{\Sigma}_u)$, we observe that

$$\begin{aligned} \mathrm{vech}(\hat{\Sigma}_u^* - \hat{\Sigma}_u) &= \mathrm{vech}[(T-p)^{-1} \sum_{t=p+1}^{T} \tilde{u}^*(t)\tilde{u}^*(t)' - (T-p)^{-1} \sum_{t=p+1}^{T} \tilde{u}(t)\tilde{u}(t)'] \\ &= \mathrm{vech}[(T-p)^{-1} \sum_{t=p+1}^{T} (\tilde{u}^*(t)\tilde{u}^*(t)' - (T-p)^{-1} \sum_{t=p+1}^{T} \tilde{u}(t)\tilde{u}(t)')]. \end{aligned}$$

Conditional on the sample and by the law of iterated expectations, we have

$$\begin{aligned} &\mathsf{E}^*[\tilde{u}^*(t)\tilde{u}^*(t)' - (T-p)^{-1} \sum_{t=p+1}^{T} \tilde{u}(t)\tilde{u}(t)'] \\ &= \mathsf{E}^*[\mathsf{E}^*[\tilde{u}^*(t)\tilde{u}^*(t)' - (T-p)^{-1} \sum_{t=p+1}^{T} \tilde{u}(t)\tilde{u}(t)' \mid \Im_t^*]] \\ &= \mathsf{E}^*[\mathsf{E}^*[\tilde{u}^*(t)\tilde{u}^*(t)' \mid \Im_t^*] - (T-p)^{-1} \sum_{t=p+1}^{T} \tilde{u}(t)\tilde{u}(t)']. \end{aligned}$$

Because

$$\mathsf{E}^*[\mathsf{E}^*[\tilde{u}^*(t)\tilde{u}^*(t)' \mid \Im_{t-1}^*] = (T-p)^{-1} \sum_{t=p+1}^{T} \mathsf{E}^*[\tilde{u}^*(t)\tilde{u}^*(t)'].$$

then

$$\mathsf{E}^*[\tilde{u}^*(t)\tilde{u}^*(t)' - (T-p)^{-1}\sum_{t=p+1}^{T}\tilde{u}(t)\tilde{u}(t)'] = 0.$$

Since

$$(T-p)^{-1}\sum_{t=p+1}^{T}(\tilde{u}^*(t)\tilde{u}^*(t)' - (T-p)^{-1}\sum_{t=p+1}^{T}\tilde{u}(t)\tilde{u}(t)')$$
$$= \mathsf{E}^*[\tilde{u}^*(t)\tilde{u}^*(t)' - (T-p)^{-1}\sum_{t=p+1}^{T}\tilde{u}(t)\tilde{u}(t)'] + o_p(1),$$

we get

$$\mathrm{vec}(\hat{\Sigma}_u^*) - \mathrm{vec}(\hat{\Sigma}_u) \underset{p}{\rightarrow} 0.$$

Similarly, we can show that

$$\mathrm{vec}(\hat{\Pi}^{c*}(k)) \underset{p}{\rightarrow} \mathrm{vec}(\hat{\Pi}^c(k)), \ \mathrm{vech}(\hat{\Sigma}_{\varepsilon,k}^*) \underset{p}{\rightarrow} \mathrm{vech}(\hat{\Sigma}_{\varepsilon,k}).$$

For $H(.)$ and $G(.)$ continuous functions, we have

$$\begin{aligned}
\ln(H(\mathrm{vec}(\hat{\Pi}^*), \mathrm{vech}(\hat{\Sigma}_u^*))) &= \ln(H(\mathrm{vec}(\hat{\Pi}), \mathrm{vech}(\hat{\Sigma}_u))) + o_p(1), \\
\ln(G(\mathrm{vec}(\hat{\Pi}^{c*}(k)), \mathrm{vech}(\hat{\Sigma}_{\varepsilon,k}^*))) &= \ln(G(\mathrm{vec}(\hat{\Pi}^c(k)), \mathrm{vech}(\hat{\Sigma}_{\varepsilon,k}))) + o_p(1).
\end{aligned}$$

By Theorems 2.1–3.4 in Paparoditis (1996) and Theorem 6 in Lewis and Reinsel (1985), we have, for $\frac{2}{3} < \delta < 1$,

$$\ln(G(\mathrm{vec}(\hat{\Pi}^{*c}(k)), \mathrm{vech}(\hat{\Sigma}_{\varepsilon,k}^*))) = \ln(G(\mathrm{vec}(\Pi^c), \mathrm{vech}(\Sigma_\varepsilon)) + O(T^{-\delta}).$$

Thus,

$$\hat{C}^*(Y \underset{h}{\longrightarrow} X \mid Z) = \ln\left(\frac{G(f(\text{vec}(\hat{\Pi}^*),\ \text{vech}(\hat{\Sigma}_u^*)))}{H(\text{vec}(\hat{\Pi}^*),\ \text{vech}(\hat{\Sigma}_u^*))}\right) + O(T^{-\delta}) + o_p(1)$$
$$= \hat{C}^*(Y \underset{h}{\longrightarrow} X \mid ZZ) + O(T^{-\delta}) + o_p(1)$$

where

$$\hat{C}^*(Y \underset{h}{\longrightarrow} X|Z) = \ln\left(\frac{G(f(\text{vec}(\hat{\Pi}^*),\ \text{vech}(\hat{\Sigma}_u^*)))}{H(\text{vec}(\hat{\Pi}^*),\ \text{vech}(\hat{\Sigma}_u^*))}\right).$$

We have also shown [see the proof of Proposition **??**] that, for $\frac{2}{3} < \delta < 1$,

$$\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z) = \ln\left(\frac{G(f(\text{vec}(\hat{\Pi}),\ \text{vech}(\hat{\Sigma}_u)))}{H(\text{vec}(\hat{\Pi}),\ \text{vech}(\hat{\Sigma}_u))}\right) + O(T^{-\delta}) + o_p(1).$$

Consequently

$$\hat{C}^*(Y \to X \mid Z) = \ln\left(\frac{G(f(\text{vec}(\hat{\Pi}),\ \text{vech}(\hat{\Sigma}_u)))}{H(\text{vec}(\hat{\Pi}),\ \text{vech}(\hat{\Sigma}_u))}\right) + O(T^{-\delta}) + o_p(1).$$

Furthermore, by Assumption 2 and a first order Taylor expansion of $\tilde{C}^*(Y \to X \mid Z)$, we have

$$\hat{C}^*(Y \underset{h}{\longrightarrow} X|Z) = \tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) + D_C \begin{pmatrix} \text{vec}(\hat{\Pi}^*) - \text{vec}(\hat{\Pi}) \\ \text{vech}(\hat{\Sigma}_u^*) - \text{vech}(\hat{\Sigma}_u) \end{pmatrix} + o_p(T^{\frac{1}{2}}).$$

and

$$T^{1/2}[\hat{C}^*(Y \underset{h}{\longrightarrow} X|Z) - \tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z)] \simeq D_C \begin{pmatrix} (T-p)^{1/2}(\text{vec}(\hat{\Pi}^*) - \text{vec}(\hat{\Pi})) \\ (T-p)^{1/2}(\text{vech}(\hat{\Sigma}_u^*) - \text{vech}(\hat{\Sigma}_u)) \end{pmatrix}.$$

By (1.40),

$$T^{1/2}[\hat{C}^*(Y \underset{h}{\longrightarrow} X|Z) - \tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z)] \xrightarrow{d} \mathcal{N}(0, \Sigma_C).$$

hence

$$T^{1/2}[\hat{C}^*(Y \underset{h}{\longrightarrow} X \mid Z) - \hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)] \xrightarrow{d} \mathcal{N}(0, \Sigma_C)$$

where

$$\Sigma_C = D_C \Omega D_C'.$$

$$\Omega = \begin{pmatrix} \Gamma^{-1} \otimes \Sigma_u & 0 \\ 0 & 2(D_3' D_3)^{-1} D_3' (\Sigma_u \otimes \Sigma_u) D_3 (D_3' D_3)^{-1} \end{pmatrix}.$$

D_3 is the duplication matrix, defined such that $\mathrm{vech}(F) = D_3 \mathrm{vech}(F)$ for any symmetric 3×3 matrix F. ■

Chapter 2

Measuring causality between volatility and returns with high-frequency data

2.1 Introduction

One of the many stylized facts about equity returns is an asymmetric relationship between returns and volatility. In the literature there are two explanations for volatility asymmetry. The first is the leverage effect. A decrease in the price of an asset increases financial leverage and the probability of bankruptcy, making the asset riskier, hence an increase in volatility [see Black (1976) and Christie (1982)]. When applied to an equity index, this original idea translates into a dynamic leverage effect.[1] The second explanation or volatility feedback effect is related to the time-varying risk premium theory: if volatility is priced, an anticipated increase in volatility would raise the rate of return, requiring an immediate stock price decline in order to allow for higher future returns [see Pindyck (1984), French, Schwert and Stambaugh (1987), Campbell and Hentschel (1992), and Bekaert and Wu (2000)].

As mentioned by Bekaert and Wu (2000) and more recently by Bollerslev et al. (2006), the difference between the leverage and volatility feedback explanations for volatility asymmetry is related to the issue of *causality.* The leverage effect explains why a negative return shock leads to higher subsequent volatility, while the volatility feedback effect justifies how an increase in volatility may result in a negative return. Thus, volatility asymmetry may result from various causality links: from returns to volatility, from volatility to returns, instantaneous causality, all of these causal effects, or just some of them.

Bollerslev et al. (2006) looked at these relationships using high frequency data and realized volatility measures. This strategy increases the chance to detect true causal links since aggregation may make the relationship between returns and volatility simultaneous. Using an observable approximation for volatility avoids the necessity to commit to a

[1]The concept of leverage effect, which means that negative returns today increases volatility of tomorrow, has been introduced in the context of the individual stocks (individual firms). However, this concept was also conserved and studied within the framework of the stock market indices by Bouchaud, Matacz, and Potters (2001), Jacquier, Polson, and Rossi (2004), Brandt and Kang (2004), Ludvigson and Ng (2005), Bollerslev et al. (2006), among others.

volatility model. Their empirical strategy is thus to use correlation between returns and realized volatility to measure and compare the magnitude of the leverage and volatility feedback effects. However, correlation is a measure of linear association but does not necessarily imply a causal relationship. In this chapter, we propose an approach which consists in modelling at high frequency both returns and volatility as a vector autoregressive (VAR) model and using short and long run causality measures proposed in chapter one to quantify and compare the strength of dynamic leverage and volatility feedback effects.

Studies focusing on the leverage hypothesis [see Christie (1982) and Schwert (1989)] conclude that it cannot completely account for changes in volatility. However, for the volatility feedback effect, there are conflicting empirical findings. French, Schwert, and Stambaugh (1987), Campbell and Hentschel (1992), and Ghysels, Santa-Clara, and Valkanov (2005), find the relation between volatility and expected returns to be positive, while Turner, Startz, and Nelson (1989), Glosten, Jagannathen, and Runkle (1993), and Nelson (1991) find the relation to be negative. Often the coefficient linking volatility to returns is statistically insignificant. Ludvigson and Ng (2005), find a strong positive contemporaneous relation between the conditional mean and conditional volatility and a strong negative lag-volatility-in-mean effect. Guo and Savickas (2006), conclude that the stock market risk-return relation is positive, as stipulated by the CAPM; however, idiosyncratic volatility is negatively related to future stock market returns. For individual assets, Bekaert and Wu (2000) argue that the volatility feedback effect dominates the leverage effect empirically. With high-frequency data, Bollerslev et al. (2006) find an important negative correlation between volatility and current and lagged returns lasting for several days. However, correlations between returns and lagged volatility are all close to zero.

A second contribution of this chapter is to show that the causality measures may help to quantify the dynamic impact of bad and good news on volatility.[2] A common approach

[2] In this study we mean by bad and good news negative and positive returns, respectively. In parallel,

for empirically visualizing the relationship between news and volatility is provided by the news-impact curve originally studied by Pagan and Schwert (1990) and Engle and Ng (1993). To study the effect of current return shocks on future expected volatility, Engle and Ng (1993) introduced the News Impact Function (hereafter NIF). The basic idea of this function is to condition at time $t+1$ on the information available at time t and earlier, and then consider the effect of the return shock at time t on volatility at time $t+1$ in isolation. Engle and Ng (1993) explained that this curve, where all the lagged conditional variances are evaluated at the level of the asset return unconditional variance, relates past positive and negative returns to current volatility. In this chapter, we propose a new curve for the impact of news on volatility based on causality measures. In contrast to the NIF of Engle and Ng (1993), our curve can be constructed for parametric and stochastic volatility models and it allows one to consider all the past information about volatility and returns. Furthermore, we build confidence intervals using a bootstrap technique around our curve, which provides an improvement over current procedures in terms of statistical inference.

Using 5-minute observations on S&P 500 Index futures contracts, we measure a weak dynamic leverage effect for the first four hours in hourly data and a strong dynamic leverage effect for the first three days in daily data. The volatility feedback effect is found to be negligible at all horizons. These findings are consistent with those in Bollerslev et al. (2006). We also use the causality measures to quantify and test statistically the dynamic impact of good and bad news on volatility. First, we assess by simulation the ability of causality measures to detect the differential effect of good and bad news in various parametric volatility models. Then, empirically, we measure a much stronger impact for bad news at several horizons. Statistically, the impact of bad news is found to be significant for the first four days, whereas the impact of good news is negligible at

there is another literature about the impact of macroeconomic news announcements on financial markets (e.g. volatility), see for example Cutler, Poterba and Summers (1989), Schwert (1981), Pearce and Roley (1985), Hardouvelis (1987), Haugen, Talmor and Torous (1991), Jain (1988), McQueen and Roley (1993), Balduzzi, Elton, and Green (2001), Andersen, Bollerslev, Diebold, and Vega (2003), and Huang (2007), among others.

all horizons.

The plan of this chapter is as follows. In section 2.2, we define volatility measures in high frequency data and we review the concept of causality at different horizons and its measures. In section 2.3, we propose and discuss VAR models that allow us to measure leverage and volatility feedback effects with high frequency data, as well as to quantify the dynamic impact of news on volatility. In section 2.4, we conduct a simulation study with several symmetric and asymmetric volatility models to assess if the proposed causality measures capture well the dynamic impact of news. Section 2.5 describes the high frequency data, the estimation procedure and the empirical findings. In section 2.6 we conclude by summarizing the main results.

2.2 Volatility and causality measures

Since we want to measure causality between volatility and returns at high frequency, we need to build measures for both volatility and causality. For volatility, we use various measures of realized volatility introduced by Andersen, Bollerslev, and Diebold (2003) [see also Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold, and Labys (2001), and Barndorff-Nielsen and Shephard (2002a,b). For causality, we rely on the short and long run causality measures proposed in chapter one.

We first set notations. We denote the time-t logarithmic price of the risky asset or portfolio by p_t and the continuously compounded returns from time t to $t+1$ by $r_{t+1} = p_{t+1} - p_t$. We assume that the price process may exhibit both stochastic volatility and jumps. It could belong to the class of continuous-time jump diffusion processes,

$$dp_t = \mu_t dt + \sigma_t dW_t + \kappa_t dq_t, \ 0 \leq t \leq T, \tag{2.1}$$

where μ_t is a continuous and locally bounded variation process, σ_t is the stochastic volatility process, W_t denotes a standard Brownian motion, dq_t is a counting process with $dq_t = 1$ corresponding to a jump at time t and $dq_t = 0$ otherwise, with jump intensity λ_t. The parameter κ_t refers to the size of the corresponding jumps. Thus, the quadratic variation of return from time t to $t+1$ is given by:

$$[r,r]_{t+1} = \int_t^{t+1} \sigma_s^2 ds + \sum_{0<s\leq t} \kappa_s^2.$$

where the first component, called integrated volatility, comes from the continuous component of (2.1), and the second term is the contribution from discrete jumps. In the absence of jumps, the second term on the right-hand-side disappears, and the quadratic variation is simply equal to the integrated volatility.

2.2.1 Volatility in high frequency data: realized volatility, bipower variation, and jumps

In this section, we define the various high-frequency measures that we will use to capture volatility. In what follows we normalize the daily time-interval to unity and we divide it into h periods. Each period has length $\Delta = 1/h$. Let the discretely sampled Δ-period returns be denoted by $r_{(t,\Delta)} = p_t - p_{t-\Delta}$ and the daily return is $r_{t+1} = \sum_{j=1}^{h} r_{(t+j.\Delta,\Delta)}$. The daily realized volatility is defined as the summation of the corresponding h high-frequency intradaily squared returns,

$$RV_{t+1} \equiv \sum_{j=1}^{h} r^2_{(t+j\Delta.\Delta)}.$$

As noted by Andersen, Bollerslev, and Diebold (2003) [see also Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold, and Labys (2001), Barndorff-Nielsen and Shephard (2002a,b) and Comte and Renault (1998)], the realized volatility satisfies

$$\lim_{\Delta \longrightarrow 0} RV_{t+1} = \int_t^{t+1} \sigma_s^2 ds + \sum_{0<s\leq t} \kappa_s^2 \tag{2.2}$$

and this means that RV_{t+1} is a consistent estimator of the sum of the integrated variance $\int_t^{t+1} \sigma_s^2 ds$ and the jump contribution.[3] Similarly, a measure of standardized bipower variation is given by

$$BV_{t+1} \equiv \frac{\pi}{2} \sum_{j=2}^{h} | r_{(t+j\Delta,\Delta)} || r_{(t+(j-1)\Delta,\Delta)} | .$$

Based on Barndorff-Nielsen and Shephard's (2003c) results [see also Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2005)], under reasonable assumptions about the dynamics of (2.1), the bipower variation satisfies,

$$\lim_{\Delta \longrightarrow 0} BV_{t+1} = \int_t^{t+1} \sigma_s^2 ds. \tag{2.3}$$

equation (2.3) means that BV_{t+1} provides a consistent estimator of the integrated variance unaffected by jumps. Finally, as noted by Barndorff-Nielsen and Shephard (2003c), combining the results in equation (2.2) and (2.3), the contribution to the quadratic variation due to the discontinuities (jumps) in the underlying price process may be consistently estimated by

$$\lim_{\Delta \longrightarrow 0} (RV_{t+1} - BV_{t+1}) = \sum_{0<s\leq t} \kappa_s^2. \tag{2.4}$$

We can also define the relative measure

$$RJ_{t+1} = \frac{(RV_{t+1} - BV_{t+1})}{RV_{t+1}} \tag{2.5}$$

or the corresponding logarithmic ratio

$$J_{t+1} = \log(RV_{t+1}) - \log(BV_{t+1}).$$

[3]See Meddahi (2002) for an interesting related literature about a theoretical comparison between integrated and realized volatility in the absence of jumps.

Huang and Tauchen (2005) argue that these are more robust measures of the contribution of jumps to total price variation.

Since in practice J_{t+1} can be negative in a given sample, we impose a non-negativity truncation of the actual empirical jump measurements,

$$J_{t+1} \equiv \max[\log(RV_{t+1}) - \log(BV_{t+1}), 0].$$

as suggested by Barndorff-Nielsen and Shephard (2003c).[4]

2.2.2 Short-run and long-run causality measures

The concept of noncausality that we consider in this chapter is defined in terms of orthogonality conditions between subspaces of a Hilbert space of random variables with finite second moments. To give a formal definition of noncausality at different horizons, we need to consider the following notations. We denote $\underline{r(t)} = \{r_{t+1-s}, s \geq 1\}$ and $\underline{\sigma_t^2} = \{\sigma_{t+1-s}^2, s \geq 1\}$ the information sets which contain all the past and present values of returns and volatility, respectively. We denote by I_t the information set which contains $\underline{r(t)}$ and $\underline{\sigma_t^2}$. For any information set B_t, we denote by $Var[r_{t+h} \mid B_t]$ (respectively $Var[\sigma_{t+h}^2 \mid B_t]$) the variance of the forecast error of r_{t+h} (respectively σ_{t+h}^2) based on the information set B_t.[5] Thus, we have the following definition of noncausality at different horizons [see Dufour and Renault (1998)].

Definition 6 *For $h \geq 1$, where h is a positive integer,*
(i) r does not cause σ^2 at horizon h given $\underline{\sigma_t^2}$, denoted $r \underset{h}{\nrightarrow} \sigma^2 \mid \underline{\sigma_t^2}$, iff

$$Var(\sigma_{t+h}^2 \mid \underline{\sigma_t^2}) = Var(\sigma_{t+h}^2 \mid I_t),$$

[4] See also Andersen, Bollerslev, and Diebold (2003).
[5] B_t can be equal to I_t, $\underline{r(t)}$, or $\underline{\sigma_t^2}$.

(ii) r does not cause σ^2 up to horizon h given $\underline{\sigma_t^2}$, denoted $r \underset{(h)}{\nrightarrow} \sigma^2 \mid \underline{\sigma_t^2}$, iff

$$r \underset{k}{\nrightarrow} \sigma^2 \mid \underline{\sigma_t^2} \text{ for } k = 1, 2, \ldots, h,$$

(iii) r does not cause σ^2 at any horizon given $\underline{\sigma_t^2}$, denoted $r \underset{(\infty)}{\nrightarrow} \sigma^2 \mid \underline{\sigma_t^2}$, iff

$$r \underset{k}{\nrightarrow} \sigma^2 \mid \underline{\sigma_t^2} \text{ for all } k = 1, 2, \ldots$$

Definition 6 corresponds to causality from r to σ^2 and means that r causes σ^2 at horizon h if the past of r improves the forecast of σ_{t+h}^2 given the information set $\underline{\sigma_t^2}$. We can similarly define noncausality at horizon h from σ^2 to r. This definition is a simplified version of the original definition given by Dufour and Renault (1998). Here we consider an information set I_t which contains only two variables of interest r and σ^2. However, Dufour and Renault (1998) [see also chapter one] consider a third variable, called an auxiliary variable, which can transmit causality between r and σ^2 at horizon h strictly higher than one even if there is no causality between the two variables at horizon 1. In the absence of an auxiliary variable, Dufour and Renault (1998) show that noncausality at horizon 1 implies noncausality at any horizon h strictly higher than one. In other words, if we suppose that $I_t = \underline{r(t)} \cup \underline{\sigma_t^2}$, then we have:

$$r \underset{1}{\nrightarrow} \sigma^2 \mid \underline{\sigma_t^2} \Longrightarrow r \underset{(\infty)}{\nrightarrow} \sigma^2 \mid \underline{\sigma_t^2}.$$

$$\sigma^2 \underset{1}{\nrightarrow} r \mid \underline{r(t)} \Longrightarrow \sigma^2 \underset{(\infty)}{\nrightarrow} r \mid \underline{r(t)}.$$

For $h \geq 1$, where h is a positive integer, a measure of causality from r to σ^2 at horizon h, denoted $C(r \underset{h}{\longrightarrow} \sigma^2)$, is given by following function [see chapter one]:

$$C(r \underset{h}{\longrightarrow} \sigma^2) = \ln \left[\frac{Var[\sigma_{t+h}^2 \mid \underline{\sigma_t^2}]}{Var[\sigma_{t+h}^2 \mid I_t]} \right].$$

Similarly, a measure of causality from σ^2 to r at horizon h, denoted $C(\sigma^2 \underset{h}{\longrightarrow} r)$, is given by:

$$C(\sigma^2 \underset{h}{\longrightarrow} r) = \ln\left[\frac{Var[r_{t+h} \mid \underline{r(t)}]}{Var[r_{t+h} \mid I_t]}\right].$$

For example, $C(r \underset{h}{\longrightarrow} \sigma^2)$ measures the causal effect from r to σ^2 at horizon h given the past of σ^2. In terms of predictability, it measures the information given by the past of r that can improve the forecast of σ^2_{t+h}. Since $Var[\sigma^2_{t+h} \mid \underline{\sigma^2_t}] \geq Var[\sigma^2_{t+h} \mid I_t]$, the function $C(r \underset{h}{\longrightarrow} \sigma^2)$ is nonnegative, as any measure must be. Furthermore, it is zero when $Var[\sigma^2_{t+h} \mid \underline{\sigma^2_t}] = Var[\sigma^2_{t+h} \mid I_t]$, or when there is no causality. However, as soon as there is causality at horizon 1, causality measures at different horizons may considerably differ.

In chapter one, a measure of instantaneous causality between r and σ^2 at horizon h is also proposed. It is given by the function

$$C(r \underset{h}{\leftrightarrow} \sigma^2) = \ln\left[\frac{Var[r_{t+h} \mid I_t]\; Var[\sigma^2_{t+h} \mid I_t]}{\det \Sigma(r_{t+h}, \sigma^2_{t+h} \mid I_t)}\right],$$

where $\det \Sigma(r_{t+h}, \sigma^2_{t+h} \mid I_t)$ represents the determinant of the variance-covariance matrix, denoted $\Sigma(r_{t+h}, \sigma^2_{t+h} \mid I_t)$, of the forecast error of the joint process $(r, \sigma^2)'$ at horizon h given the information set I_t. Finally, in chapter one we propose a measure of dependence between r and σ^2 at horizon h which is given by the following formula:

$$C^{(h)}(r, \sigma^2) = \ln\left[\frac{Var[r_{t+h} \mid \underline{r(t)}]\,Var[\sigma^2_{t+h} \mid \underline{\sigma^2_t}]}{\det \Sigma(r_{t+h}, \sigma^2_{t+h} \mid I_t)}\right].$$

This last measure can be decomposed as follows:

$$C^{(h)}(r, \sigma^2) = C(r \underset{h}{\longrightarrow} \sigma^2) + C(\sigma^2 \underset{h}{\longrightarrow} r) + C(r \underset{h}{\leftrightarrow} \sigma^2). \tag{2.6}$$

2.3 Measuring causality in a VAR model

In this section, we first study the relationship between the return r_{t+1} and its volatility σ_{t+1}^2. Our objective is to measure and compare the strength of *dynamic* leverage and volatility feedback effects in high-frequency equity data. These effects are quantified within the context of an autoregressive (VAR) linear model and by using short and long run causality measures proposed in chapter one. Since the volatility asymmetry may be the result of causality from returns to volatility [leverage effect], from volatility to returns [volatility feedback effect], instantaneous causality, all of these causal effects, or some of them, this section aims at measuring all these effects and to compare them in order to determine the most important one. We also measure the dynamic impact of return news on volatility where we differentiate good and bad news.

2.3.1 Measuring the leverage and volatility feedback effects

We suppose that the joint process of returns and logarithmic volatility, $(r_{t+1}, \ln(\sigma_{t+1}^2))'$ follows an autoregressive linear model

$$\begin{pmatrix} r_{t+1} \\ \ln(\sigma_{t+1}^2) \end{pmatrix} = \mu + \sum_{j=1}^{p} \Phi_j \begin{pmatrix} r_{t+1-j} \\ \ln(\sigma_{t+1-j}^2) \end{pmatrix} + u_{t+1}. \tag{2.7}$$

where

$$\mu = \begin{pmatrix} \mu_r \\ \mu_{\sigma^2} \end{pmatrix}, \; \Phi_j = \begin{bmatrix} \Phi_{11,j} & \Phi_{12,j} \\ \Phi_{21,j} & \Phi_{22,j} \end{bmatrix} \text{ for } j = 1, \ldots, p, \; u_{t+1} = \begin{pmatrix} u_{t+1}^r \\ u_{t+1}^{\sigma^2} \end{pmatrix},$$

$$E[u_t] = 0 \text{ and } E\left[u_t u_t'\right] = \begin{cases} \Sigma_u & \text{for } s = t \\ 0 & \text{for } s \neq t \end{cases}.$$

In the empirical application σ_{t+1}^2 will be replaced by the realized volatility RV_{t+1} or the bipower variation BV_{t+1}. The disturbance u_{t+1}^r is the one-step-ahead error when r_{t+1} is forecast from its own past and the past of $\ln(\sigma_{t+1}^2)$, and similarly $u_{t+1}^{\sigma^2}$ is the one-step-ahead error when $\ln(\sigma_{t+1}^2)$ is forecast from its own past and the past of r_{t+1}. We suppose that these disturbances are each serially uncorrelated, but may be correlated with each other contemporaneously and at various leads and lags. Since u_{t+1}^r is uncorrelated with I_t,[6] the equation for r_{t+1} represents the linear projection of r_{t+1} on I_t. Likewise, the equation for $\ln(\sigma_{t+1}^2)$ represents the linear projection of $\ln(\sigma_{t+1}^2)$ on I_t.

Equation (2.7) allows one to model the first two conditional moments of the asset returns. We model conditional volatility as an exponential function process to guarantee that it is positive. The first equation of the $VAR(p)$ in (2.7) describes the dynamics of the return as

$$r_{t+1} = \mu_r + \sum_{j=1}^{p} \Phi_{11,j} r_{t+1-j} + \sum_{j=1}^{p} \Phi_{12,j} \ln(\sigma_{t+1-j}^2) + u_{t+1}^r. \tag{2.8}$$

This equation allows to capture the temporary component of Fama and French's (1988) permanent and temporary components model, in which stock prices are governed by a random walk and a stationary autoregressive process, respectively. For $\Phi_{12,j} = 0$, this model of the temporary component is the same as that of Lamoureux and Zhou (1996); see Brandt and Kang (2004), and Whitelaw (1994). The second equation of $VAR(p)$ describes the volatility dynamics as

$$\ln(\sigma_{t+1}^2) = \mu_{\sigma^2} + \sum_{j=1}^{p} \Phi_{21,j} r_{t+1-j} + \sum_{j=1}^{p} \Phi_{22,j} \ln(\sigma_{t+1-j}^2) + u_{t+1}^{\sigma^2}, \tag{2.9}$$

and it represents the standard stochastic volatility model. For $\Phi_{21,j} = 0$, equation (2.9) can be viewed as the stochastic volatility model estimated by Wiggins (1987), Andersen and Sorensen (1994), and many others. However, in this chapter we consider that σ_{t+1}^2 is not a latent variable and it can be approximated by realized or bipower variations from high-frequency data. We also note that the conditional mean equation includes the volatility-in-mean model used by French et al. (1987) and Glosten et al. (1993) to explore the contemporaneous relationship between the conditional mean and volatility

[6] $I_t = \{r_{t+1-s}, s \geq 1\} \cup \{\sigma_{t+1-s}^2, s \geq 1\}$.

[see Brandt and Kang (2004)]. To illustrate the connection to the volatility-in-mean model, we pre-multiply the system in (2.7) by the matrix

$$\begin{bmatrix} 1 & -\frac{Cov(r_{t+1}, \ln(\sigma^2_{t+1}))}{Var[\ln(\sigma^2_{t+1})|I_t]} \\ -\frac{Cov(r_{t+1}, \ln(\sigma^2_{t+1}))}{Var[r_{t+1}|I_t]} & 1 \end{bmatrix}.$$

Then, the first equation of r_{t+1} is a linear function of $\underline{r(t)}$, $\underline{\ln(\sigma^2_{t+1})}$, and the disturbance $u^r_{t+1} - \frac{Cov(r_{t+1}, \ln(\sigma^2_{t+1}))}{Var[\ln(\sigma^2_{t+1})|I_t]} u^{\sigma^2}_{t+1}$.[7] Since this disturbance is uncorrelated with $u^{\sigma^2}_{t+1}$, it is uncorrelated with $\ln(\sigma^2_{t+1})$ as well as with $\underline{r(t)}$ and $\underline{\ln(\sigma^2_t)}$. Hence the linear projection of r_{t+1} on $\underline{r(t)}$ and $\underline{\ln(\sigma^2_{t+1})}$,

$$r_{t+1} = \nu_r + \sum_{j=1}^{p} \phi_{11,j} r_{t+1-j} + \sum_{j=0}^{p} \phi_{12,j} \ln(\sigma^2_{t+1-j}) + u^r_{t+1} \tag{2.10}$$

is provided by the first equation of the new system. The parameters ν_r, $\phi_{11,j}$, and $\phi_{12,j}$, for $j = 0, 1, ..., p$, are functions of parameters in the vector μ and matrix Φ_j, for $j = 1, ..., p$. Equation (2.10) is a generalized version of the usual volatility-in-mean model, in which the conditional mean depends contemporaneously on the conditional volatility. Similarly, the existence of the linear projection of $\ln(\sigma^2_{t+1})$ on $\underline{r(t+1)}$ and $\underline{\ln(\sigma^2_t)}$,

$$\ln(\sigma^2_{t+1}) = \nu_{\sigma^2} + \sum_{j=0}^{p} \phi_{21,j} r_{t+1-j} + \sum_{j=1}^{p} \phi_{22,j} \ln(\sigma^2_{t+1-j}) + u^{\sigma^2}_{t+1} \tag{2.11}$$

follows from the second equation of the new system. The parameters ν_{σ^2}, $\phi_{21,j}$, and $\phi_{22,j}$, for $j = 1, ..., p$, are functions of parameters in the vector μ and matrix Φ_j, for $j = 1, ..., p$. The volatility model given by equation (2.11) captures the persistence of volatility through the terms $\phi_{22,j}$. In addition, it incorporates the effects of the mean on volatility, both at the contemporaneous and intertemporal levels through the coefficients $\phi_{21,j}$, for $j = 0, 1, ..., p$.

Let us now consider the matrix

[7] $\underline{\ln(\sigma^2_{t+1})} = \{\ln(\sigma^2_{t+2-s}), s \geq 1\}$.

$$\Sigma_u = \begin{bmatrix} \Sigma_{u^r} & C' \\ C & \Sigma_{u^{\sigma^2}} \end{bmatrix}.$$

where Σ_{u^r} and $\Sigma_{u^{\sigma^2}}$ represent the variances of the one-step-ahead forecast errors of return and volatility, respectively. C represents the covariance between these errors. Based on equation (2.7), the forecast error of $(r_{t+h}, \ln(\sigma^2_{t+h}))'$ is given by:

$$e\left[(r_{t+h}, \ln(\sigma^2_{t+h}))'\right] = \sum_{i=0}^{h-1} \psi_i u_{t+h-i}, \tag{2.12}$$

where the coefficients ψ_i, for $i = 0, ..., h-1$, represent the impulse response coefficients of the $MA(\infty)$ representation of model (2.7). These coefficients are given by the following equations

$$\begin{aligned}
&\psi_0 = I, \\
&\psi_1 = \Phi_1\psi_0 = \Phi_1, \\
&\psi_2 = \Phi_1\psi_1 + \Phi_2\psi_0 = \Phi_1^2 + \Phi_2, \\
&\psi_3 = \Phi_1\psi_2 + \Phi_2\psi_1 + \Phi_2\psi_0 = \Phi_1^3 + \Phi_1\Phi_2 + \Phi_2\Phi_1 + \Phi_3, \\
&\dots
\end{aligned} \tag{2.13}$$

where I is a 2×2 identity matrix and

$$\Phi_j = 0, \text{ for } j \geq p+1.$$

The covariance matrix of the forecast error (2.12) is given by

$$Var[e[(r_{t+h}, \ln(\sigma^2_{t+h}))']] = \sum_{i=0}^{h-1} \psi_i \Sigma_u \psi_i'. \tag{2.14}$$

We also consider the following restricted model:

$$\begin{pmatrix} r_{t+1} \\ \ln(\sigma^2_{t+1}) \end{pmatrix} = \bar{\mu} + \sum_{j=1}^{\bar{p}} \bar{\Phi}_j \begin{pmatrix} r_{t+1-j} \\ \ln(\sigma^2_{t+1-j}) \end{pmatrix} + \bar{u}_{t+1} \tag{2.15}$$

where

$$\bar{\Phi}_j = \begin{bmatrix} \bar{\Phi}_{11,j} & 0 \\ 0 & \bar{\Phi}_{22,j} \end{bmatrix} \text{ for } j = 1, ..., \bar{p}, \tag{2.16}$$

$$\bar{\mu} = \begin{pmatrix} \bar{\mu}_r \\ \bar{\mu}_{\sigma^2} \end{pmatrix}, \quad \bar{u}_{t+1} = \begin{pmatrix} \bar{u}^r_{t+1} \\ \bar{u}^{\sigma^2}_{t+1} \end{pmatrix},$$

$$E\left[\bar{u}_t\right] = 0 \text{ and } E\left[\bar{u}_t \bar{u}_t^{'}\right] = \begin{cases} \Sigma_{\bar{u}} & \text{for } s = t \\ 0 & \text{for } s \neq t \end{cases}.$$

$$\Sigma_{\bar{u}} = \begin{bmatrix} \Sigma_{\bar{u}^r} & \bar{C} \\ \bar{C} & \Sigma_{\bar{u}^{\sigma^2}} \end{bmatrix}.$$

Zero values in $\bar{\Phi}_j$ mean that there is noncausality at horizon 1 from returns to volatility and from volatility to returns. As mentioned in subsection 2.2.2, in a bivariate system, noncausality at horizon one implies noncausality at any horizon h strictly higher than one. This means that the absence of leverage effects at horizon one (respectively the absence of volatility feedback effects at horizon one) which corresponds to $\bar{\Phi}_{21,j} = 0$, for $j = 1, ..., \bar{p}$, (respectively $\bar{\Phi}_{12,j} = 0$, for $j = 1, ..., \bar{p}$,) is equivalent to the absence of leverage effects (respectively volatility feedback effects) at any horizon $h \geq 1$.

To compare the forecast error variance of model (2.7) with that of model (2.15), we assume that $p = \bar{p}$. Based on the restricted model (2.15), the covariance matrix of the forecast error of $(r_{t+h}, \ln(\sigma^2_{t+h}))'$ is given by:

$$\overline{Var}[t + h \mid t] = \sum_{i=0}^{h-1} \bar{\psi}_i \Sigma_{\bar{u}} \bar{\psi}_i^{'}, \tag{2.17}$$

where the coefficients $\bar{\psi}_i$, for $i = 0, ..., h - 1$, represent the impulse response coefficients of the $MA(\infty)$ representation of model (2.15). They can be calculated in the same way as in (2.13). From the covariance matrices (2.14) and (2.17), we define the following measures of leverage and volatility feedback effects at any horizon h, where $h \geq 1$,

$$C(r \underset{h}{\longrightarrow} \ln(\sigma^2)) = \ln\left[\frac{\sum_{i=0}^{h-1} e_2'(\psi_i \Sigma_{\bar{u}} \psi_i')e_2}{\sum_{i=0}^{h-1} e_2'(\psi_i \Sigma_u \psi_i')e_2}\right], \; e_2 = (0, 1)', \tag{2.18}$$

$$C(\ln(\sigma^2) \underset{h}{\longrightarrow} r) = \ln\left[\frac{\sum_{i=0}^{h-1} e_1'(\psi_i \Sigma_{\bar{u}} \psi_i')e_1}{\sum_{i=0}^{h-1} e_1'(\psi_i \Sigma_u \psi_i')e_1}\right], \; e_1 = (1, 0)'. \tag{2.19}$$

The parametric measure of instantaneous causality at horizon h, where $h \geq 1$, is given by the following function

$$C(r \underset{h}{\leftrightarrow} \ln(\sigma^2)) = \ln\left[\frac{(\sum_{i=0}^{h-1} e_2'(\psi_i \Sigma_u \psi_i')e_2)\,(\sum_{i=0}^{h-1} e_1'(\psi_i \Sigma_u \psi_i')e_1)}{\det(\sum_{i=0}^{h-1} \psi_i \Sigma_u \psi_i')}\right].$$

Finally, the parametric measure of dependence at horizon h can be deduced from its decomposition given by equation (2.6).

2.3.2 Measuring the dynamic impact of news on volatility

In what follows we study the *dynamic* impact of bad news (negative innovations in returns) and good news (positive innovations in returns) on volatility. We quantify and compare the strength of these effects in order to determine the most important ones. To analyze the impact of news on volatility, we consider the following model,

$$\begin{aligned} \ln(\sigma_{t+1}^2) &= \mu_\sigma + \sum_{j=1}^{p} \varphi_j^\sigma \ln(\sigma_{t+1-j}^2) + \sum_{j=1}^{p} \varphi_j^- [r_{t+1-j} - E_{t-j}(r_{t+1-j})]^- \\ &\quad + \sum_{j=1}^{p} \varphi_j^+ [r_{t+1-j} - E_{t-j}(r_{t+1-j})]^+ + u_{t+1}^\sigma. \end{aligned} \tag{2.20}$$

where for $j = 1, ..., p$,

$$[r_{t+1-j} - E_{t-j}(r_{t+1-j})]^{-} = \begin{cases} r_{t+1-j} - E_{t-j}(r_{t+1-j}), \text{ if } r_{t+1-j} - E_{t-j}(r_{t+1-j}) \leq 0, \\ \\ 0, \text{ otherwise}, \end{cases} \tag{2.21}$$

$$[r_{t+1-j} - E_{t-j}(r_{t+1-j})]^{+} = \begin{cases} r_{t+1-j} - E_{t-j}(r_{t+1-j}), \text{ if } r_{t+1-j} - E_{t-j}(r_{t+1-j}) \geq 0, \\ \\ 0, \text{ otherwise}, \end{cases} \tag{2.22}$$

with

$$E\left[u_{t+1}^{\sigma}\right] = 0 \text{ and } E\left[(u_{t+1}^{\sigma})^2\right] = \begin{cases} \Sigma_{u^\sigma} & \text{for } s = t \\ 0 & \text{for } s \neq t \end{cases}.$$

Equation (2.20) represents the linear projection of volatility on its own past and the past of centered negative and positive returns. This regression model allows one to capture the effect of centered negative or positive returns on volatility through the coefficients φ_j^- or φ_j^+ respectively, for $j = 1, ..., p$. It also allows one to examine the different effects that large and small negative and/or positive information shocks have on volatility. This will provide a check on the results obtained in the literature on GARCH modeling, which has put forward overwhelming evidence on the effect of negative shocks on volatility.

Again, in our empirical applications, σ_{t+1}^2 will be replaced by realized volatility RV_{t+1} or bipower variation BV_{t+1}. Furthermore, the conditional mean return will be approximated by the following rolling-sample average:

$$\hat{E}_t(r_{t+1}) = \frac{1}{m} \sum_{j=1}^{m} r_{t+1-j}.$$

where we take an average around $m = 15, 30, 90, 120,$ and 240 days. Now, let us consider the following restricted models:

$$\ln(\sigma_{t+1}^2) = \theta_\sigma + \sum_{i=1}^{\bar{p}} \bar{\varphi}_i^\sigma \ln(\sigma_{t+1-i}^2) + \sum_{i=1}^{\bar{p}} \bar{\varphi}_i^+ [r_{t+1-i} - E_{t-j}(r_{t+1-i})]^+ + e_{t+1}^\sigma \tag{2.23}$$

$$\ln(\sigma_{t+1}^2) = \dot{\theta}_\sigma + \sum_{i=1}^{\dot{p}} \dot{\varphi}_i^\sigma \ln(\sigma_{t+1-i}^2) + \sum_{i=1}^{\dot{p}} \dot{\varphi}_i^- [r_{t+1-j} - E_{t-j}(r_{t+1-j})]^- + v_{t+1}^\sigma. \tag{2.24}$$

Equation (2.23) represents the linear projection of volatility $\ln(\sigma_{t+1}^2)$ on its own past and the past of positive returns. Similarly, equation (2.24) represents the linear projection of volatility $\ln(\sigma_{t+1}^2)$ on its own past and the past of centred negative returns.

In our empirical application we also consider a model with non centered negative and positive returns:

$$\ln(\sigma_{t+1}^2) = \omega_\sigma + \sum_{j=1}^{p} \phi_j^\sigma \ln(\sigma_{t+1-j}^2) + \sum_{j=1}^{p} \phi_j^- r_{t+1-j}^- + \sum_{j=1}^{p} \phi_j^+ r_{t+1-j}^+ + \epsilon_{t+1}^\sigma,$$

where for $j = 1, \ldots, p$,

$$r_{t+1-j}^- = \begin{cases} r_{t+1-j}, \text{ if } r_{t+1-j} \leq 0 \\ 0, \text{ otherwise,} \end{cases}$$

$$r_{t+1-j}^+ = \begin{cases} r_{t+1-j}, \text{ if } r_{t+1-j} \geq 0 \\ 0, \text{ otherwise,} \end{cases}$$

$$E\left[\epsilon^{\sigma}_{t+1}\right] = 0 \text{ and } E\left[(\epsilon^{\sigma}_{t+1})^2\right] = \begin{cases} \Sigma_{\epsilon^{\sigma}} & \text{for } s = t \\ 0 & \text{for } s \neq t \end{cases},$$

and the corresponding restricted volatility models:

$$\ln(\sigma^2_{t+1}) = \lambda_{\sigma} + \sum_{i=1}^{\bar{p}} \bar{\phi}^{\sigma}_i \ln(\sigma^2_{t+1-i}) + \sum_{i=1}^{\bar{p}} \bar{\phi}^{+}_i r^{+}_{t+1-i} + v^{\sigma}_{t+1} \tag{2.25}$$

$$\ln(\sigma^2_{t+1}) = \bar{\lambda}_{\sigma} + \sum_{i=1}^{\dot{p}} \dot{\phi}^{\sigma}_i \ln(\sigma^2_{t+1-i}) + \sum_{i=1}^{\dot{p}} \dot{\phi}^{-}_i r^{-}_{t+1-j} + \varepsilon^{\sigma}_{t+1}. \tag{2.26}$$

To compare the forecast error variances of model (2.20) with those of models (2.23) and (2.24), we assume that $p = \bar{p} = \dot{p}$. Thus, a measure of the impact of bad news on volatility at horizon h, where $h \geq 1$, is given by the following equation:

$$C(r^{-} \underset{h}{\rightarrow} \ln(\sigma^2)) = \ln\left[\frac{Var[e^{\sigma}_{t+h} \mid \underline{\ln(\sigma^2_t)}, \underline{r^{+}(t)}]}{Var[u^{\sigma}_{t+h} \mid J_t]}\right].$$

Similarly, a measure of the impact of good news on volatility at horizon h is given by:

$$C(r^{+} \underset{h}{\rightarrow} \ln(\sigma^2)) = \ln\left[\frac{Var[v^{\sigma}_{t+h} \mid \underline{\ln(\sigma^2_t)}, \underline{r^{-}(t)}]}{Var[u^{\sigma}_{t+h} \mid J_t]}\right]$$

where

$$\underline{r^{-}(t)} = \{[r_{t-s} - E_{t-1-s}(r_{t-s})]^{-}, s \geq 0\},$$

$$\underline{r^{+}(t)} = \{[r_{t-s} - E_{t-1-s}(r_{t-s})]^{+}, s \geq 0\},$$

$$J_t = \underline{\ln(\sigma^2(t))} \cup \underline{r^{-}(t)} \cup \underline{r^{+}(t)}.$$

We also define a function which allows us to compare the impact of bad and good news on volatility. This function can be defined as follows:

$$C(r^-/r^+ \underset{h}{\rightarrow} \ln(\sigma^2)) = \ln\left[\frac{Var[e^{\sigma}_{t+h} \mid \underline{\ln(\sigma^2_t)}, \underline{r^+(t)}]}{Var[v^{\sigma}_{t+h} \mid \underline{\ln(\sigma^2_t)}, \underline{r^-(t)}]}\right].$$

When $C(r^-/r^+ \underset{h}{\rightarrow} \ln(\sigma^2)) \geq 0$, this means that bad news have more impact on volatility than good news. Otherwise, good news will have more impact on volatility than bad news.

2.4 A simulation study

In this section we verify with a thorough simulation study the ability of the causality measures to detect the well-documented asymmetry in the impact of bad and good news on volatility [see Pagan and Schwert (1990), Gouriéroux and Monfort (1992), and Engle and Ng (1993)]. To assess the asymmetry in leverage effect, we consider the following structure. First, we suppose that returns are governed by the process

$$r_{t+1} = \sqrt{\sigma_t}\varepsilon_{t+1} \tag{2.27}$$

where $\varepsilon_{t+1} \sim \mathcal{N}(0,1)$ and σ_t represents the conditional volatility of return r_{t+1}. Since we are only interested in studying the asymmetry in leverage effect, equation (2.27) does not allow for a volatility feedback effect. Second, we assume that σ_t follows one of the following heteroskedastic forms:

1. $GARCH(1,1)$ model:

$$\sigma_t = \omega + \beta\sigma_{t-1} + \alpha\varepsilon^2_{t-1}; \tag{2.28}$$

2. $EGARCH(1,1)$ model:

$$\log(\sigma_t) = \omega + \beta\log(\sigma_{t-1}) + \gamma\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}}} + \alpha\left[\frac{\mid \varepsilon_{t-1} \mid}{\sqrt{\sigma_{t-1}}} - \sqrt{2/\pi}\right]; \tag{2.29}$$

3. Nonlinear $NL\text{-}GARCH(1,1)$ model:

$$\sigma_t = \omega + \beta\sigma_{t-1} + \alpha \mid \varepsilon_{t-1} \mid^{\gamma}; \tag{2.30}$$

4. $GJR\text{-}GARCH(1,1)$ model:

$$\sigma_t = \omega + \beta\sigma_{t-1} + \alpha\varepsilon_{t-1}^2 + \gamma I_{t-1}\varepsilon_{t-1}^2 \tag{2.31}$$

where

$$I_{t-1} = \begin{cases} 1, & \text{if } \varepsilon_{t-1} \leq 0 \\ 0, & \text{otherwise} \end{cases} \quad \text{for } t = 1, ..., T \; ;$$

5. Asymmetric $AGARCH(1,1)$ model:

$$\sigma_t = \omega + \beta\sigma_{t-1} + \alpha(\varepsilon_{t-1} + \gamma)^2; \tag{2.32}$$

6. $VGARCH(1,1)$ model:

$$\sigma_t = \omega + \beta\sigma_{t-1} + \alpha(\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}}} + \gamma)^2; \tag{2.33}$$

7. Nonlinear Asymmetric $GARCH(1,1)$ model or $NGARCH(1,1)$:

$$\sigma_t = \omega + \beta\sigma_{t-1} + \alpha(\varepsilon_{t-1} + \gamma\sqrt{\sigma_{t-1}})^2. \tag{2.34}$$

$GARCH$ and $NL\text{-}GARCH$ models are, by construction, symmetric. Thus, we expect that the curves of causality measures for bad and good news will be the same. Similarly, because $EGARCH$, $GJR\text{-}GARCH$, $AGARCH$, $VGARCH$, and $NGARCH$ are asymmetric we expect that these curves will be different.

Our simulation study consists in simulating returns from equation (2.27) and volatilities from one of the models given by equations (2.28)-(2.34). Once return and volatilities are simulated, we use the model described in subsection 2.3.2 to evaluate the causality measures of bad and good news for each of the above parametric models. All simulated samples are of size $n = 40,000$. We consider a large sample to eliminate the uncertainty in the estimated parameters. The parameter values for the different parametric models considered in our simulations, are reported in Table 1. [8]

In figures 1-9 we report the impact of bad and good news on volatility for the various volatility models, in the order shown above. For the $NL\text{-}GARCH(1,1)$ model we select three values for γ : 0.5, 1.5, 2.5. Two main conclusions can be drawn form these figures. First, from figures 1, 4, 5, and 6 we see that $GARCH$ and $NL\text{-}GARCH$ are symmetric: the bad and good news have the same impact on volatility. Second, in figures 2, 3, 7, 8, and 9, we observe that $EGARCH$, $GJR\text{-}GARCH$, $AGARCH$, $VGARCH$, and $NGARCH$ are asymmetric: the bad and good news have different impact curves. More particularly, bad news have more impact on volatility than good news.

Considering the parameter values given by Table 1 in Appendix [see Engle and Ng (1993)],[9] we found that the above parametric volatility models provide different responses to bad and good news. In the presence of bad news, Figure 10 shows that the magnitude of the volatility response is the most important in the $NGARCH$ model, followed by the $AGARCH$ and $GJR\text{-}GARCH$ models. The effect is negligible in $EGARCH$ and $VGARCH$ models. The impact of good news on volatility is more observable in $AGARCH$ and $NGARCH$ models [see Figure 11]. Overall, we can conclude that the causality measures capture quite well the effects of returns on volatility both qualitatively and quantitatively. We now apply these measures to actual data. Instead of estimating a model for volatility as most of the previous studies have done [see for example Campbell and Hentschel (1992), Bekaert and Wu (2000), Glosten, Jagannathen, and Runkle (1993), and Nelson (1991)], we use a proxy measure given by realized volatility or bipower variation based on high-frequency data.

[8] We also consider other parameter values from a paper by Engle and Ng (1993). The corresponding results are available upon request from the authors. These results are similar to those shown in this paper.

[9] These parameters are the results of an estimation of different parametric volatility models using the daily returns series of the Japanese TOPIX index from January 1, 1980 to December 31, 1988 [see Engle and Ng (1993) for more details].

2.5 An empirical application

In this section, we first describe the data used to measure causality in the VAR models of the previous sections. Then we explain how to estimate confidence intervals of causality measures for leverage and volatility feedback effects. Finally, we discuss our findings.

2.5.1 Data

Our data consists of high-frequency tick-by-tick transaction prices for the S&P 500 Index futures contracts traded on the Chicago Mercantile Exchange, over the period January 1988 to December 2005 for a total of 4494 trading days. We eliminated a few days where trading was thin and the market was open for a shortened session. Due to the unusually high volatility at the opening, we also omit the first five minutes of each trading day [see Bollerslev et al. (2006)]. For reasons associated with microstructure effects we follow Bollerslev et al. (2006) and the literature in general and aggregate returns over five-minute intervals. We calculate the continuously compounded returns over each five-minute interval by taking the difference between the logarithm of the two tick prices immediately preceding each five-minute mark to obtain a total of 77 observations per day [see Dacorogna et al. (2001) and Bollerslev et al. (2006) for more details]. We also construct hourly and daily returns by summing 11 and 77 successive five-minute returns, respectively.

Summary statistics for the five-minute, hourly, and daily returns are given in Table 2. The daily returns are displayed in Figure 16. Looking at Table 2 and Figure 16 we can state three main stylized facts. First, the unconditional distributions of the five-minute, hourly, and daily returns show the expected excess kurtosis and negative skewness. The sample kurtosis is much greater than the normal value of three for all three series. Second, whereas the unconditional distribution of the hourly returns appears to be skewed to the left, the sample skewness coefficients for the five-minute and daily returns are, loosely speaking, both close to zero.

We also compute various measures of return volatility, namely realized volatility and bipower variation, both in levels and in logarithms. The time series plots [see Figures 17, 18, 19, and 20] show clearly the familiar volatility clustering effect, along with a few occasional very large absolute returns. It also follows from Table 3 that the unconditional distributions of realized and bipower volatility measures are highly skewed and leptokurtic. However, the logarithmic transform renders both measures approximately normal [Andersen, Bollerslev, Diebold, and Ebens (2001)]. We also note that the descriptive statistics for the relative jump measure, J_{t+1}, clearly indicate a positively skewed and leptokurtic distribution.

One way to test if realized and bipower volatility measures are significantly different is to test for the presence of jumps in the data. We recall that,

$$\lim_{\Delta \to 0}(RV_{t+1}) = \int_t^{t+1} \sigma_s^2 ds + \sum_{0<s\leq t} \kappa_s^2, \tag{2.35}$$

where $\sum_{0<s\leq t} \kappa_s^2$ represents the contribution of jumps to total price variation. In the absence of jumps, the second term on the right-hand-side disappears, and the quadratic variation is simply equal to the integrated volatility: or asymptotically ($\Delta \to 0$) the realized variance is equal to the bipower variance.

Many statistics have been proposed to test for the presence of jumps in financial data [see for example Barndorff-Nielsen and Shephard (2003b), Andersen, Bollerslev, and Diebold (2003), Huang and Tauchen (2005), among others]. In this chapter, we test for the presence of jumps in our data by considering the following test statistics:

$$z_{QP,l,t} = \frac{RV_{t+1} - BV_{t+1}}{\sqrt{((\frac{\pi}{2})^2 + \pi - 5)\Delta QP_{t+1}}}, \tag{2.36}$$

$$z_{QP,t} = \frac{\log(RV_{t+1}) - \log(BV_{t+1})}{\sqrt{((\frac{\pi}{2})^2 + \pi - 5)\Delta \frac{QP_{t+1}}{BV_{t+1}^2}}}, \tag{2.37}$$

$$z_{QP,lm,t} = \frac{\log(RV_{t+1}) - \log(BV_{t+1})}{\sqrt{((\frac{\pi}{2})^2 + \pi - 5)\Delta \max(1, \frac{QP_{t+1}}{BV_{t+1}^2})}}. \tag{2.38}$$

where QP_{t+1} is the realized Quad-Power Quarticity [Barndorff-Nielsen and Shephard (2003a)], with

$$QP_{t+1} = h\mu_1^{-4} \sum_{j=4}^{h} | r_{(t+j.\Delta,\Delta)} || r_{(t+(j-1).\Delta,\Delta)} || r_{(t+(j-2).\Delta,\Delta)} || r_{(t+(j-3).\Delta,\Delta)} |.$$

and $\mu_1 = \sqrt{\frac{2}{\pi}}$. For each time t, the statistics $z_{QP,l,t}$, $z_{QP,t}$, and $z_{QP,lm,t}$ follow a Normal distribution $\mathcal{N}(0,1)$ as $\Delta \to 0$, under the assumption of no jumps. The results of testing for jumps in our data are plotted in Figures 12-15. Figure 12 represents the Quantile-Quantile Plot (hereafter QQ Plot) of the relative measure of jumps given by equation (2.5). The other Figures, see 13, 14, and 15, represent the QQ Plots of the $z_{QP,l,t}$, $z_{QP,t}$, and $z_{QP,lm,t}$ statistics, respectively. When there are no jumps, we expect that the blue and red lines in Figures 12-15 will coincide. However, as these figures show, the two lines are clearly distinct, indicating the presence of jumps in our data. Therefore, we will present our results for both realized volatility and bipower variation.

2.5.2 Estimation of causality measures

We apply short-run and long-run causality measures to quantify the strength of relationships between return and volatility. We use OLS to estimate the $VAR(p)$ models described in sections 2.3 and 2.3.2 and the Akaike information criterion to specify their orders. To obtain consistent estimates of the causality measures we simply replace the unknown parameters by their estimates.[10] We calculate causality measures for various horizons $h = 1, ..., 20$. A higher value for a causality measure indicates a stronger causality. We also compute the corresponding nominal 95% bootstrap confidence intervals

[10] See proof of consistency of the estimation in chapter one.

according to the procedure described in the Appendix.

2.5.3 Results

We examine several empirical issues regarding the relationship between volatility and returns. Because volatility is unobservable and high-frequencies data were not available, these issues have been addressed before mainly in the context of volatility models. Recently, Bollerslev et al. (2006) looked at these relationships using high frequency data and realized volatility measures. As they emphasize, the fundamental difference between the leverage and the volatility feedback explanations lies in the direction of causality. The leverage effect explains why a low return causes higher subsequent volatility, while the volatility feedback effect captures how an increase in volatility may cause a negative return. However, they studied only correlations between returns and volatility at various leads and lags and not causality relationships between both. The concept of causality introduced by Granger (1969) necessitates an information set and is conducted in the framework of a model between the variables of interest. Moreover, it is also important economically to measure the strength of this causal link and to test if the effect is significantly different from zero. In measuring causal relationship, aggregation is of course a major problem. Low frequency data may mask the true causal relationship between the variables. Looking at high-frequency data offers an ideal setting to isolate, if any, causal effects. Formulating a VAR model to study causality allows also to distinguish between the immediate or current effects between the variables and the effects of the lags of one variable on the other. It should be emphasized also that even for studying the relationship at daily frequencies, using high-frequency data to construct daily returns and volatilities provides better estimates than using daily returns as most previous studies have done. Since realized volatility is an approximation of the true unobservable volatility we study the robustness of the results to another measure, the bipower variation, which is robust to the presence of jumps.

Our empirical results will be presented mainly through graphs. Each figure will report the causality measure on the vertical axis while the horizon will appear on the horizontal axis. We also draw in each figure the 95% bootstrap confidence intervals. With five-minute intervals we could conceivably estimate the VAR model at this frequency. However if we wanted to allow for enough time for the effects to develop we would need a large number of lags in the VAR model and sacrifice efficiency in the estimation. This problem arises in studies of volatility forecasting. Researchers have use several schemes to group five-minute intervals, in particular the HAR-RV or the MIDAS schemes.[11]

We decided to aggregate the returns at hourly frequency and study the corresponding intradaily causality relationship between returns and volatility. As illustrated in figures 24 (log realized volatility) and 25 (log bipower variation), we find that the leverage effect is statistically significant for the first four hours but that it is small in magnitude. The volatility feedback effect in hourly data is negligible at all horizons [see tables 6 and 7].

Using daily observations, calculated with high frequency data, we measure a strong leverage effect for the first three days. This result is the same with both realized and bipower variations [see figures 22 and 23]. The volatility feedback effect is found to be negligible at all horizons [see tables 4 and 5]. By comparing these two effects, we find that the leverage effect is more important than the volatility feedback effect [see figures 30 and 31]. The comparison between the leverage effects in hourly and daily data reveal that this effect is more important in daily than in hourly returns [see figures 32 and 33].

If the feedback effect from volatility to returns is almost-non-existent, it is apparent in figures 26 and 27 that the instantaneous causality between these variables exists and remains economically and statistically important for several days. This means that volatility has a contemporaneous effect on returns, and similarly returns have a contemporaneous effect on volatility. These results are confirmed with both realized and bipower variations. Furthermore, as illustrated in figures 28 and 29, dependence between

[11]The HAR-RV scheme, in which the realized volatility is parameterized as a linear function of the lagged realized volatilities over different horizons has been proposed by Müller et al. (1997) and Corsi (2003). The MIDAS scheme, based on the idea of distributed lags, has been analyzed and estimated by Ghysels, Santa-Clara and Valkanov (2002).

volatility and returns is also economically and statistically important for several days.

Since only the causality from returns to volatility is significant, it is important to check if negative and positive returns have a different impact on volatility. To answer this question we have calculated the causality measures from centered and non centered positive and negative returns to volatility. The empirical results are graphed in figures 34-45 and reported in tables 8-11. We find a much stronger impact of bad news on volatility for several days. Statistically, the impact of bad news is significant for the first four days, whereas the impact of good news is negligible at all horizons. Figures 46 and 47 make it possible to compare for both realized and bipower variations the impact of bad and good news on volatility. As we can see, bad news have more impact on volatility than good news at all horizons.

Finally, to study the temporal aggregation effect on the relationship between returns and volatility, we compare the conditional dependence between returns and volatility at several levels of aggregation: one hour, one day, two days, 3 days, 6 days, 14 days, and 21 days. The empirical results show that the dependence between returns and volatility is an increasing function of temporal aggregation [see Figure 50]. This is still true for the 21 first days, after which the dependence decreases.

2.6 Conclusion

In this chapter we analyze and quantify the relationship between volatility and returns with high-frequency equity returns. Within the framework of a vector autoregressive linear model of returns and realized volatility or bipower variation, we quantify the dynamic leverage and volatility feedback effects by applying short-run and long-run causality measures proposed in chapter one. These causality measures go beyond simple correlation measures used recently by Bollerslev, Litvinova, and Tauchen (2006).

Using 5-minute observations on S&P 500 Index futures contracts, we measure a weak dynamic leverage effect for the first four hours in hourly data and a strong dynamic leverage effect for the first three days in daily data. The volatility feedback effect is found to be negligible at all horizons

We also use causality measures to quantify and test statistically the dynamic impact of good and bad news on volatility. First, we assess by simulation the ability of causality measures to detect the differential effect of good and bad news in various parametric volatility models. Then, empirically, we measure a much stronger impact for bad news at several horizons. Statistically, the impact of bad news is significant for the first four days, whereas the impact of good news is negligible at all horizons.

2.7 Appendix: bootstrap confidence intervals of causality measures

We compute the nominal 95% bootstrap confidence intervals of the causality measures as follows [see chapter one]:

(1) Estimate by OLS the $VAR(p)$ process given by equation (2.15) and save the residuals

$$\hat{u}(t) = \begin{pmatrix} r_t \\ RV_t \end{pmatrix} - \hat{\mu} - \sum_{j=1}^{p} \hat{\Phi}_j \begin{pmatrix} r_{t-j} \\ \ln(RV_{t-j}) \end{pmatrix}, \text{ for } t = p+1, ..., T,$$

where $\hat{\mu}$ and $\hat{\Phi}_j$ are the OLS regression estimates of μ and Φ_j, for $j = 1, ..., p$.

(2) Generate $(T-p)$ bootstrap residuals $\hat{u}^*(t)$ by random sampling with replacement from the residuals $\hat{u}(t)$, $t = p+1, ..., T$.

(3) Generate a random draw for the vector of p initial observations

$$w(0) = ((r_1, \ln((RV_1))', ..., (r_p, \ln(RV_p))')'.$$

(4) Given $\hat{\mu}$ and $\hat{\Phi}_j$, for $j = 1, ..., p$, $\hat{u}^*(t)$, and $w(0)$, generate bootstrap data for the dependent variable $(r_t^*, \ln(RV_t)^*)'$ from equation:

$$\begin{pmatrix} r_t^* \\ \ln(RV_t)^* \end{pmatrix} = \hat{\mu} + \sum_{j=1}^{p} \hat{\Phi}_j \begin{pmatrix} r_{t-j}^* \\ \ln(RV_{t-j})^* \end{pmatrix} + \hat{u}^*(t), \text{ for } t = p+1, ..., T.$$

(5) Calculate the bootstrap OLS regression estimates

$$\hat{\Phi}^* = (\hat{\mu}^*, \hat{\Phi}_1^*, \hat{\Phi}_2^*, ..., \hat{\Phi}_p^*) = \hat{\Gamma}^{*-1}\hat{\Gamma}_1^*,$$

$$\hat{\Sigma}_u^* = \sum_{t=p+1}^{T} \hat{u}^*(t)\hat{u}^*(t)'/(T-p),$$

where $\hat{\Gamma}^* = (T-p)^{-1}\sum_{t=p+1}^{T} w^*(t)w^*(t)'$, for $w^*(t) = ((r_t^*, \ln(RV_t)^*)', ..., (r_{t-p+1}^*, \ln(RV_{t-p+1})^*)')'$,

$\hat{\Gamma}_1^* = (T-p)^{-1}\sum_{t=p+1}^{T} w^*(t)(r_{t+1}^*, \ln(RV_{t+1})^*)'$, and

$$\hat{u}^*(t) = \begin{pmatrix} r_t^* \\ \ln(RV_t)^* \end{pmatrix} - \hat{\mu} - \sum_{j=1}^{p} \hat{\Phi}_j \begin{pmatrix} r_{t-j}^* \\ \ln(RV_{t-j})^* \end{pmatrix}.$$

(6) Estimate the constrained model of $\ln(RV_t)$ or r_t using the bootstrap sample $\{(r_t^*, \ln(RV_t)^*)'\}_{t=1}^{T}$.

(7) Calculate the causality measures at horizon h, denoted $\hat{C}^{(j)*}(r \underset{h}{\longrightarrow} \ln(RV))$ and $\hat{C}^{(j)*}(\ln(RV) \underset{h}{\longrightarrow} r)$, using equations (2.18) and (2.19), respectively.

(8) Choose B such $\frac{1}{2}\alpha(B+1)$ is an integer and repeat steps (2) – (7) B times.[12]

(9) Finally, calculate the α and 1-α percentile interval endpoints of the distributions of $\hat{C}^{(j)*}(r \underset{h}{\longrightarrow} \ln(RV))$ and $\hat{C}^{(j)*}(\ln(RV) \underset{h}{\longrightarrow} r)$.

A proof of the asymptotic validity of the bootstrap confidence intervals of the causality measures is provided in chapter one.

[12]Where 1-α is the considered level of confidence interval.

Table 1: Parameter values of different $GARCH$ models

	ω	β	α	γ
$GARCH$	2.7910^{-5}	0.86695	0.093928	–
$EGARCH$	−0.290306	0.97	0.093928	−0.09
$NL\text{-}GARCH$	2.7910^{-5}	0.86695	0.093928	0.5, 1.5, 2.5
$GJR\text{-}GARCH$	2.7910^{-5}	0.8805	0.032262	0.10542
$AGARCH$	2.7910^{-5}	0.86695	0.093928	−0.1108
$VGARCH$	2.7910^{-5}	0.86695	0.093928	−0.1108
$NGARCH$	2.7910^{-5}	0.86695	0.093928	−0.1108

Note: The table summarizes the parameter values for parametric volatility models considered in our simulations study.

Table 2: Summary statistics for S&P 500 futures returns, 1988-2005.

$Variables$	$Mean$	$St.Dev.$	$Median$	$Skewness$	$Kurtosis$
$Five$ − minute	$6.9505e-006$	0.000978	$0.00e-007$	−0.0818	73.9998
$Hourly$	$1.3176e-005$	0.0031	$0.00e-007$	−0.4559	16.6031
$Daily$	$1.4668e-004$	0.0089	$1.1126e-004$	−0.1628	12.3714

Note: The table summarizes the Five-minute, Hourly, and Daily returns distributions for the S&P 500 index contracts. The sample covers the period from 1988 to December 2005 for a total of 4494 trading days.

Table 3: Summary statistics for daily volatilities, 1988-2005.

$Variables$	$Mean$	$St.Dev.$	$Median$	$Skewness$	$Kurtosis$
RV_t	$8.1354e-005$	$1.2032e-004$	$4.9797e-005$	8.1881	120.7530
BV_t	$7.6250e-005$	$1.0957e-004$	$4.6956e-005$	6.8789	78.9491
$\ln(RV_t)$	−9.8582	0.8762	−9.9076	0.4250	3.3382
$\ln(BV_t)$	−9.9275	0.8839	−9.9663	0.4151	3.2841
J_{t+1}	0.0870	0.1005	0.0575	1.6630	7.3867

Note: The table summarizes the Daily volatilities distributions for the S&P 500 index contracts. The sample covers the period from 1988 to December 2005 for a total of 4494 trading days.

Table 4: Causality Measure of Daily Feedback Effect: $ln(RV)$

$C(\ln(RV) \underset{h}{\rightarrow} r)$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
Point estimate	0.0019	0.0019	0.0019	0.0011
95% Bootstrap interval	[0.0007, 0.0068]	[0.0005, 0.0065]	[0.0004, 0.0061]	[0.0002, 0.0042]

Table 5: Causality Measure of Daily Feedback Effect: $ln(BV)$

$C(\ln(BV) \underset{h}{\rightarrow} r)$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
Point estimate	0.0017	0.0017	0.0016	0.0011
95% Bootstrap interval	[0.0007, 0.0061]	[0.0005, 0.0056]	[0.0004, 0.0055]	[0.0002, 0.0042]

Table 6: Causality Measure of Hourly Feedback Effect: $ln(RV)$

$C(\ln(RV) \underset{h}{\rightarrow} r)$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
Point estimate	0.00016	0.00014	0.00012	0.00012
95% Bootstrap interval	[0.0000, 0.0007]	[0.0000, 0.0006]	[0.0000, 0.0005]	[0.0000, 0.0005]

Table 7: Causality Measure of Hourly Feedback Effect: $ln(BV)$

$C(\ln(BV) \underset{h}{\rightarrow} r)$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
Point estimate	0.00022	0.00020	0.00019	0.00015
95% Bootstrap interval	[0.0000, 0.0008]	[0.0000, 0.0007]	[0.0000, 0.0007]	[0.0000, 0.0005]

Note: Tables 4-7 summarize the estimation results of causality measures from daily realized volatility to daily returns, daily bipower variation to daily returns, hourly realized volatility to hourly returns, and hourly bipower variation to hourly returns, respectively. The second row in each table gives the point estimate of the causality measures at $h = 1, .., 4$. The third row gives the 95% corresponding percentile bootstrap interval.

Table 8: Measuring the impact of good news on volatility: Centered positive returns, $ln(RV)$
$C([r_{t+1-j} - E_{t-j}(r_{t+1-j})]^{+} \underset{h}{\rightarrow} \ln(RV))$

$\widehat{E_t(r_{t+1})} = \frac{1}{15}\sum_{j=1}^{15} r_{t+1-j}$

	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.00076	0.00075	0.00070	0.00041
95% Bootstrap interval	$[0.0003, 0.0043]$	$[0.0002, 0.0039]$	$[0.0001, 0.0034]$	$[0, 0.0030]$

$\widehat{E_t(r_{t+1})} = \frac{1}{30}\sum_{j=1}^{30} r_{t+1-j}$

	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.00102	0.00071	0.00079	0.00057
95% Bootstrap interval	$[0.00047, 0.00513]$	$[0.00032, 0.00391]$	$[0.00031, 0.00362]$	$[0, 0.00321]$

$\widehat{E_t(r_{t+1})} = \frac{1}{90}\sum_{j=1}^{90} r_{t+1-j}$

	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.0013	0.00087	0.00085	0.00085
95% Bootstrap interval	$[0.0004, 0.0059]$	$[0.00032, 0.0044]$	$[0.0002, 0.0041]$	$[0.0001, 0.0039]$

$\widehat{E_t(r_{t+1})} = \frac{1}{120}\sum_{j=1}^{120} r_{t+1-j}$

	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.0011	0.00076	0.00072	0.00074
95% Bootstrap interval	$[0.0004, 0.0054]$	$[0.00029, 0.0041]$	$[0.00024, 0.00386]$	$[0, 0.00388]$

$\widehat{E_t(r_{t+1})} = \frac{1}{240}\sum_{j=1}^{240} r_{t+1-j}$

	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.0011	0.00069	0.00067	0.0007
95% Bootstrap interval	$[0.0004, 0.0053]$	$[0.0003, 0.0041]$	$[0.0002, 0.0035]$	$[0, 0.0034]$

Note: The table summarizes the estimation results of causality measures from centered positive returns to realized volatility under five estimators of the average returns. In each of the five small tables, the second row gives the point estimate of the causality measures at $h = 1, ...4$. The third row gives the 95% corresponding percentile bootstrap interval.

Table 9: Measuring the impact of good news on volatility: Centred positive returns, $ln(BV)$

$$C([r_{t+1-j} - E_{t-j}(r_{t+1-j})]^{+} \underset{h}{\rightarrow} \ln(RV))$$

$$\widehat{E_t(r_{t+1})} = \frac{1}{15}\sum_{j=1}^{15} r_{t+1-j}$$

	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.0008	0.0008	0.00068	0.00062
95% Bootstrap interval	$[0.00038, 0.0045]$	$[0.00029, 0.0041]$	$[0.00021, 0.0035]$	$[0, 0.0034]$

$$\widehat{E_t(r_{t+1})} = \frac{1}{30}\sum_{j=1}^{30} r_{t+1-j}$$

	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.0012	0.00076	0.00070	0.00072
95% Bootstrap interval	$[0.0005, 0.0053]$	$[0.0003, 0.0041]$	$[0.0002, 0.0039]$	$[0.0001, 0.0038]$

$$\widehat{E_t(r_{t+1})} = \frac{1}{90}\sum_{j=1}^{90} r_{t+1-j}$$

	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.0018	0.0009	0.0008	0.0010
95% Bootstrap interval	$[0.0006, 0.0065]$	$[0.0003, 0.0044]$	$[0.0002, 0.0041]$	$[0.0001, 0.0042]$

$$\widehat{E_t(r_{t+1})} = \frac{1}{120}\sum_{j=1}^{120} r_{t+1-j}$$

	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.0016	0.0008	0.0007	0.0009
95% Bootstrap interval	$[0.0006, 0.0063]$	$[0.00026, 0.0047]$	$[0.0002, 0.0042]$	$[0.0001, 0.0044]$

$$\widehat{E_t(r_{t+1})} = \frac{1}{240}\sum_{j=1}^{240} r_{t+1-j}$$

	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.0015	0.0007	0.0006	0.0008
95% Bootstrap interval	$[0.0005, 0.0057]$	$[0.00029, 0.0044]$	$[0.00020, 0.0038]$	$[0.0001, 0.0037]$

Note: The table summarizes the estimation results of causality measures from centered positive returns to bipower variation under five estimators of the average returns. In each of the five small tables, the second row gives the point estimate of the causality measures at $h = 1, ..., 4$. The third row gives the 95% corresponding percentile bootstrap interval.

Table 10: Measuring the impact of good news on volatility: Noncentered positive returns, $ln(RV)$

$C(r^+ \underset{h}{\rightarrow} \ln(RV))$	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.0027	0.0012	0.0008	0.0009
95% Bootstrap interval	[0.0011, 0.0077]	[0.0004, 0.0048]	[0.0002, 0.0041]	[0.0001, 0.0038]

Table 11: Measuring the impact of good news on volatility: Noncentered positive returns, $ln(BV)$

$C(r^+ \underset{h}{\rightarrow} \ln(BV))$	$h=1$	$h=2$	$h=3$	$h=4$
Point estimate	0.0035	0.0013	0.0008	0.0010
95% Bootstrap interval	[0.0016, 0.0087]	[0.0004, 0.0051]	[0.0002, 0.0039]	[0.0001, 0.0043]

Note: Tables 10-11 summarize the estimation results of causality measures from noncentered positive returns to realized volatility and noncentered positive returns to bipower variation, respectively. The second row in each table gives the point estimate of the causality measures at $h = 1, .., 4$. The third row gives the 95% corresponding percentile bootstrap interval.

Figure1: Impact of bad and good news in GARCH(1,1)

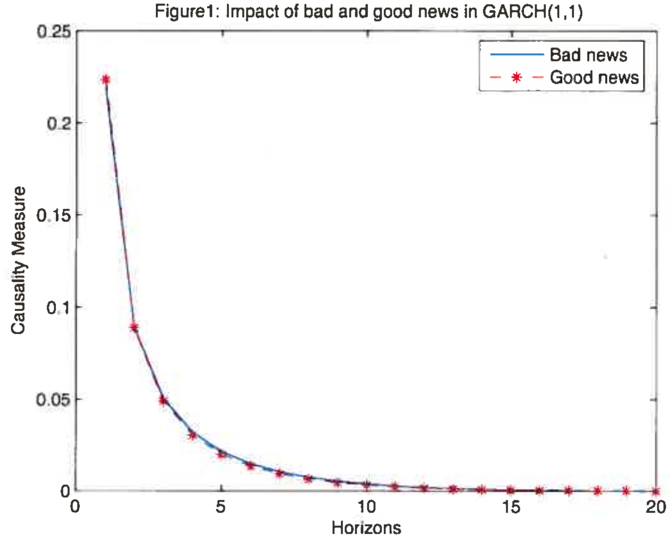


Figure 2: Impact of bad and good news in EGARCH(1,1) model

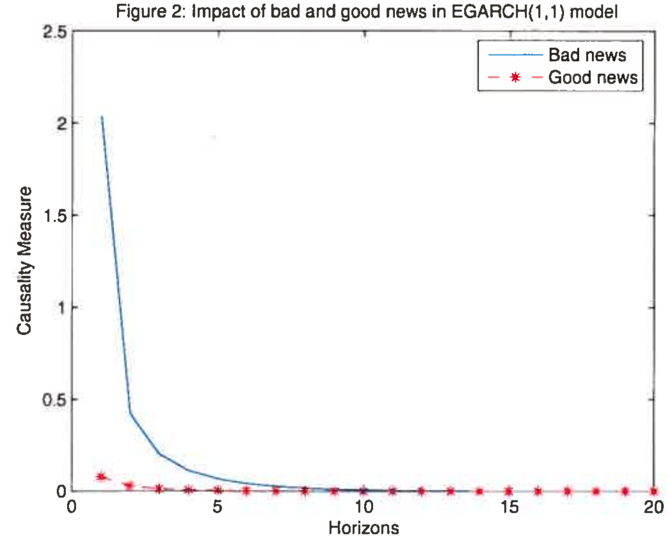


Figure 3: Impact of bad and good news in GJR-GARCH(1,1) model

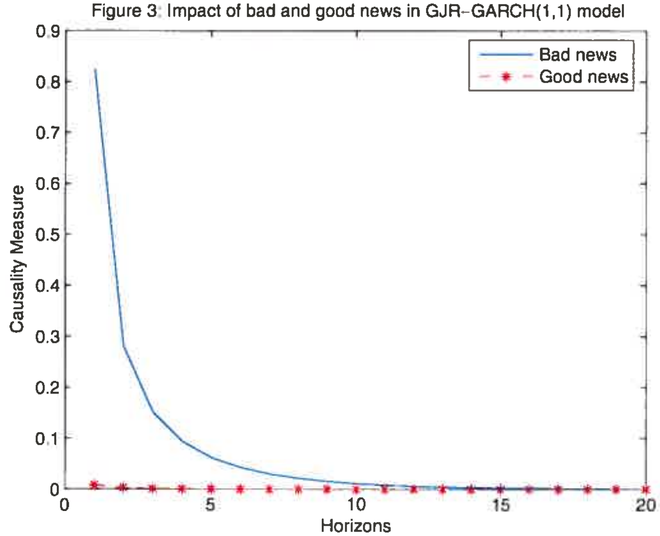


Figure 4: Impact of bad and good news in NL-GARCH(1,1) model with lamda=0.5

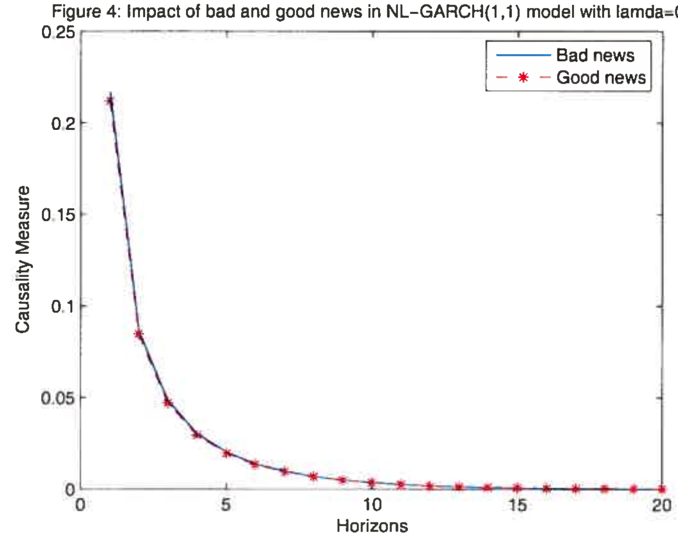


Figure 5: Impact of bad and good news in NL-GARCH(1,1) model with lamda=1

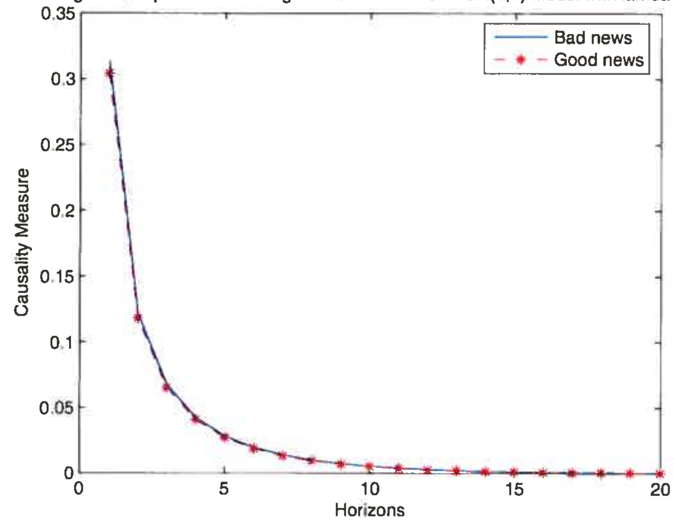


Figure 6: Impact of bad and good news in NL-GARCH(1,1) model with lamda=1.5

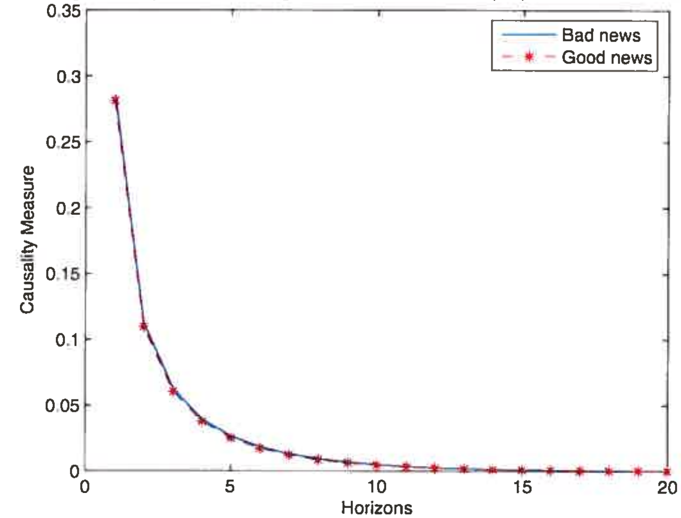


Figure 7: Measuring the impact of bad and good news in AGARCH(1,1)

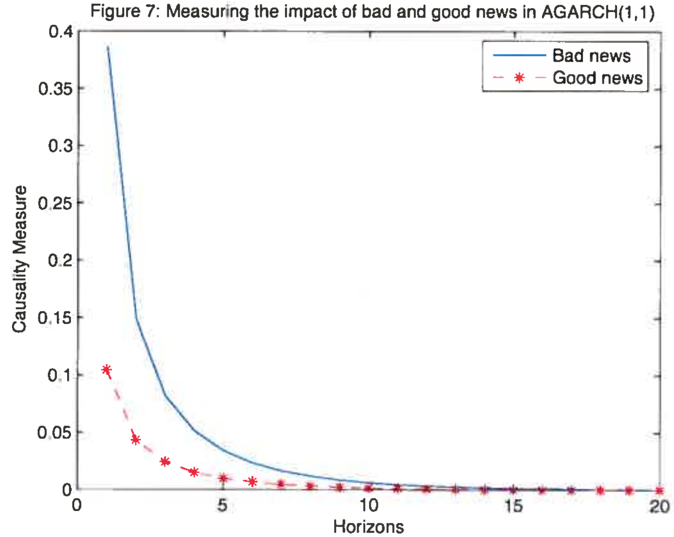


Figure 8: Impact of bad and good news in VGARCH(1,1) model

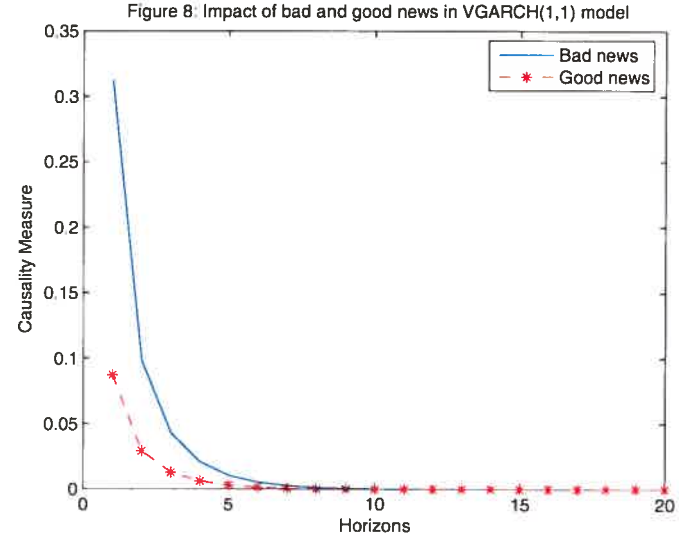


Figure 9: Impact of bad and good news in NGARCH(1,1) model

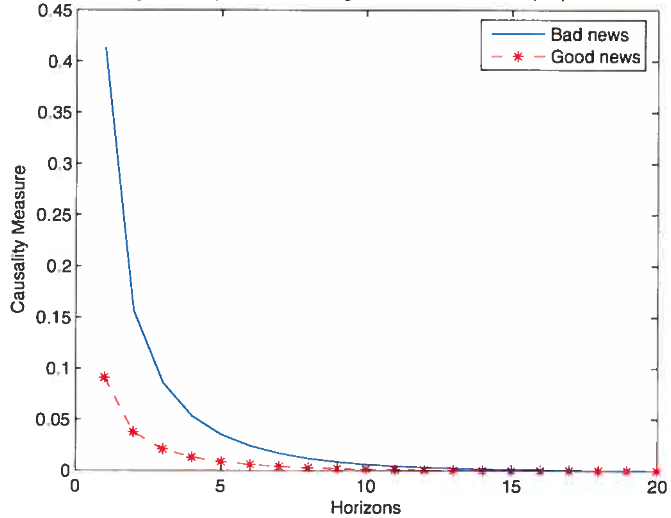


Figure 10: Response of volatility to bad news in different asymmetric GARCH models

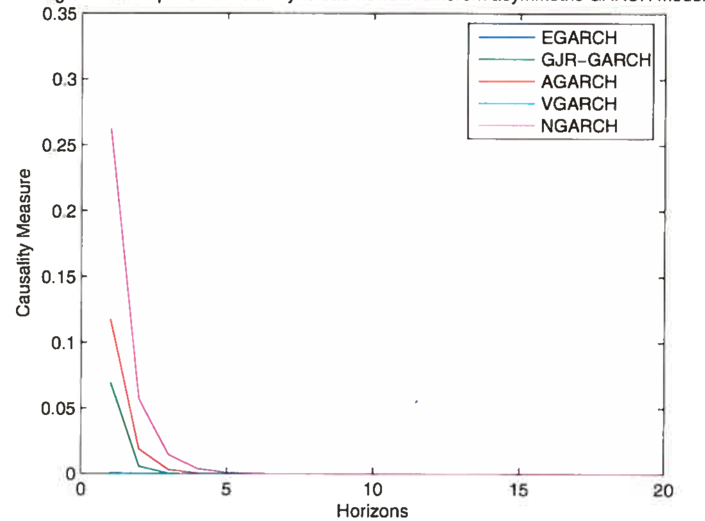


Figure 11: Response of volatility to good news in different asymmetric GARCH models

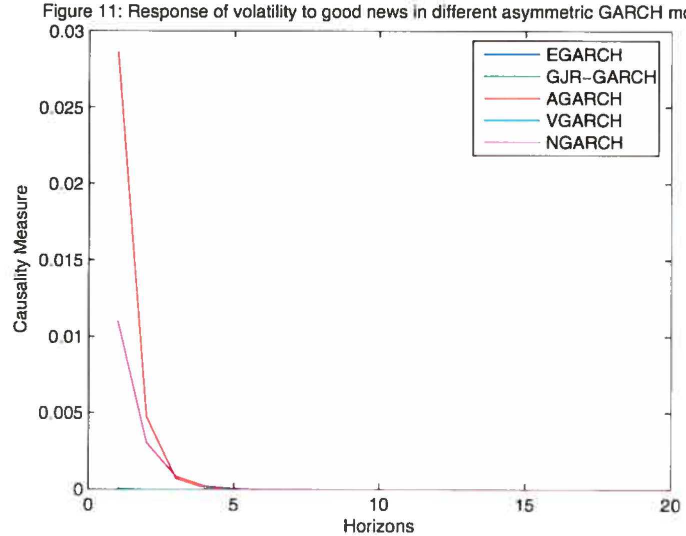


Figure 12: QQ Plot of relative jump measure versus Standard Normal

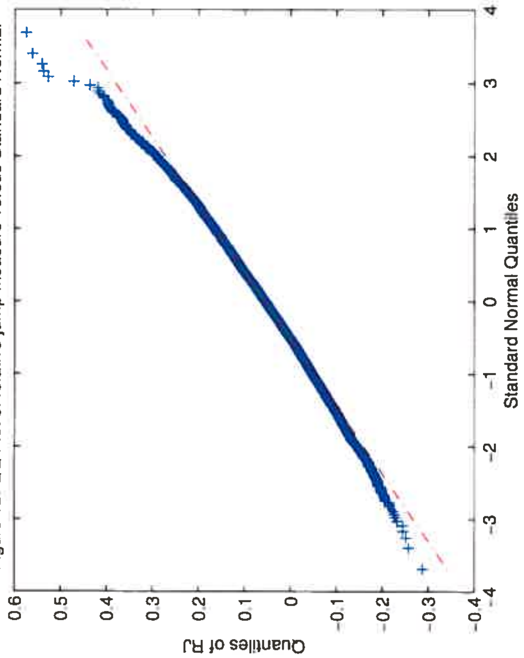


Figure 13: QQ Plot of zQP versus Standard Normal

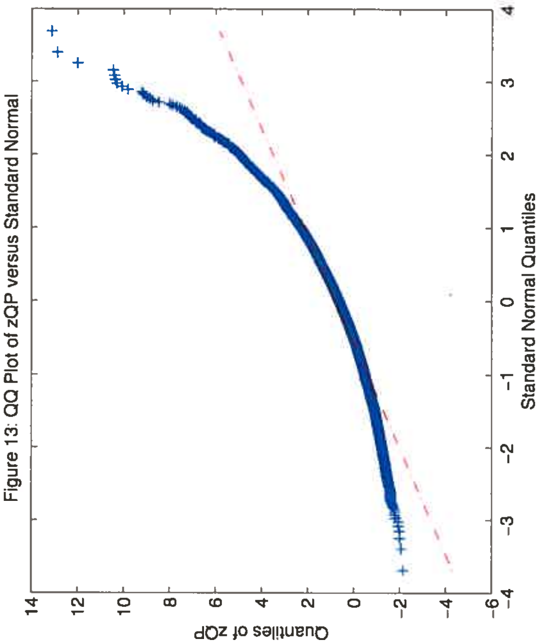


Figure 14: QQ Plot of zQPl versus Standard Normal

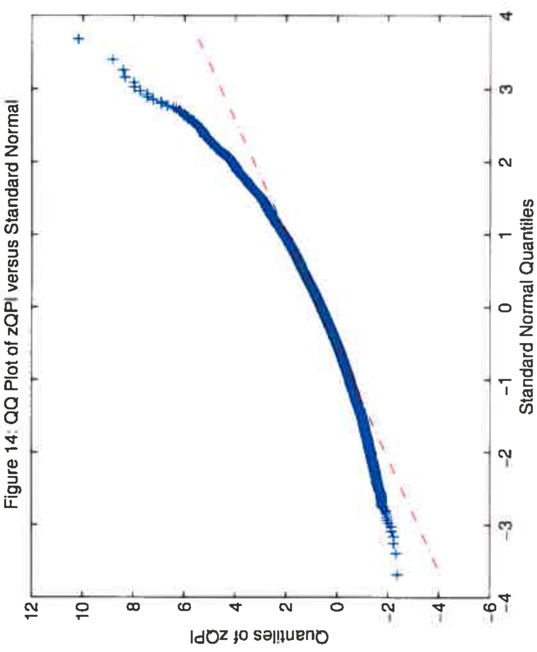


Figure 15: QQ Plot of zQPm versus Standard Normal

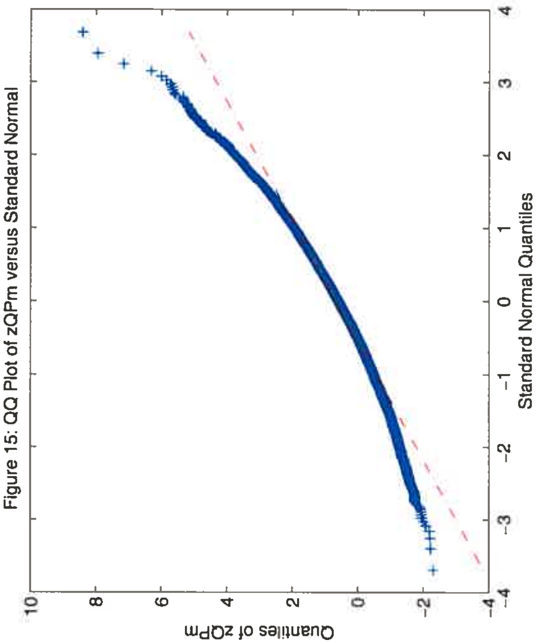
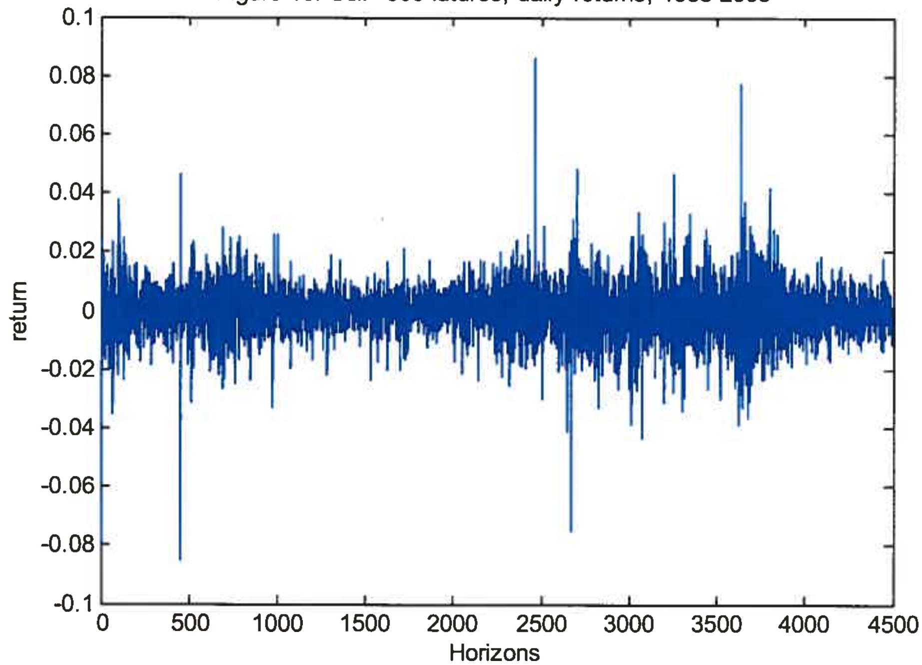


Figure 16: S&P 500 futures, daily returns, 1988-2005



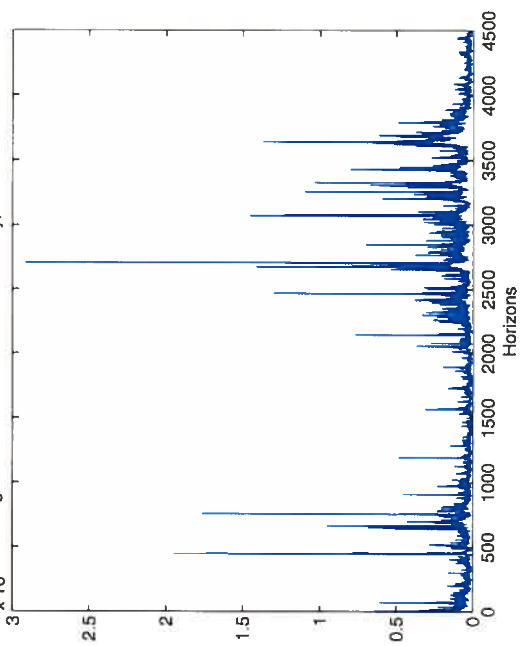


Figure 17: S&P 500 Realized volatility, 1988–2005

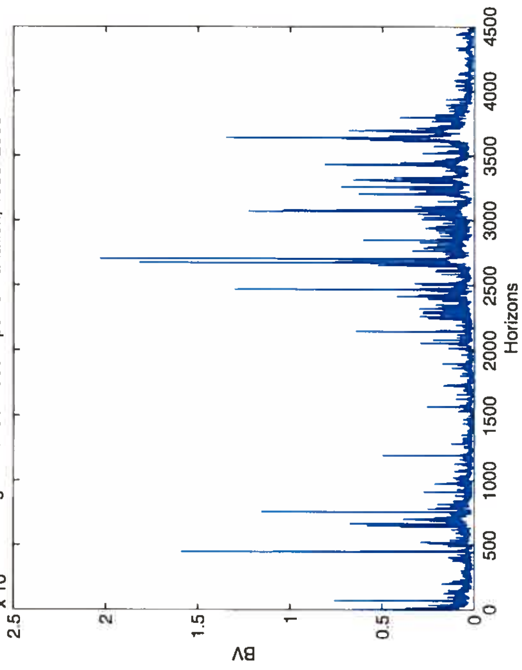


Figure 18: S&P 500 Bipower variation, 1988–2005

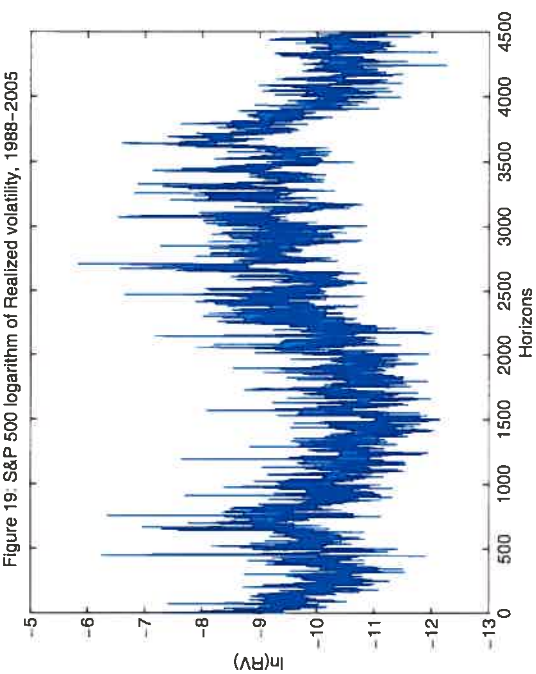


Figure 19: S&P 500 logarithm of Realized volatility, 1988–2005

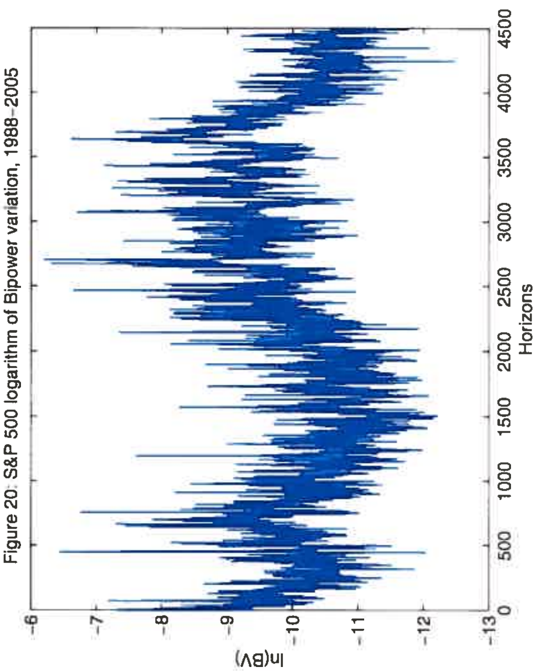


Figure 20: S&P 500 logarithm of Bipower variation, 1988–2005

Figure 21: S&P 500 Jumps, 1988-2005

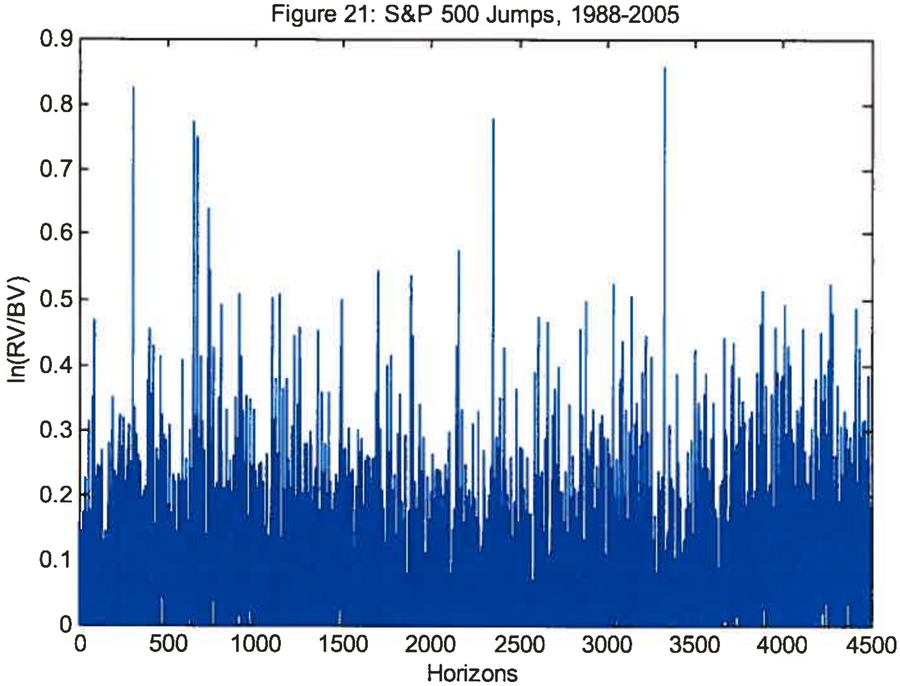


Figure 22: Causality Measures for Leverage Effect (ln(RV))

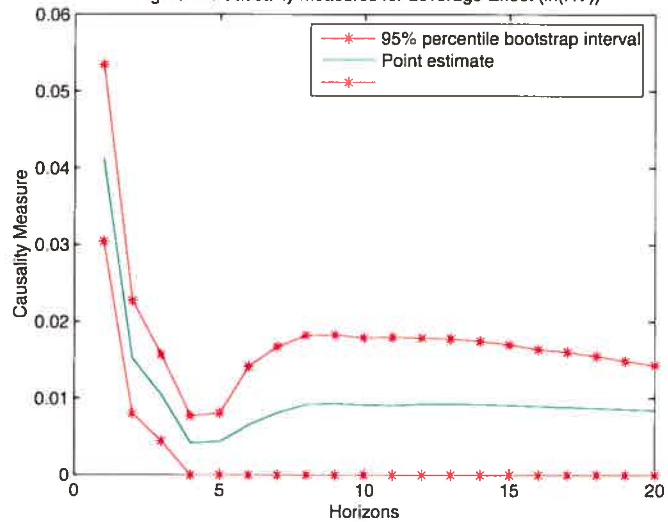


Figure 23: Causality Measures for Daily Leverage Effect (ln(BV))

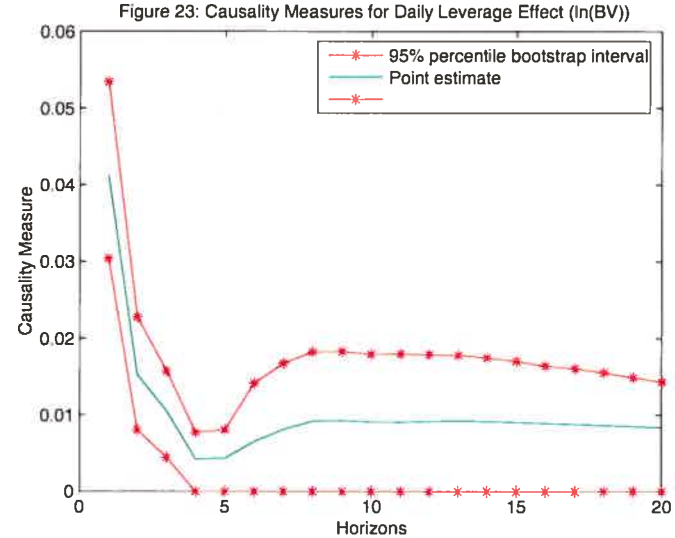


Figure 24: Causality Measures for Hourly Leverage Effect (ln(RV))

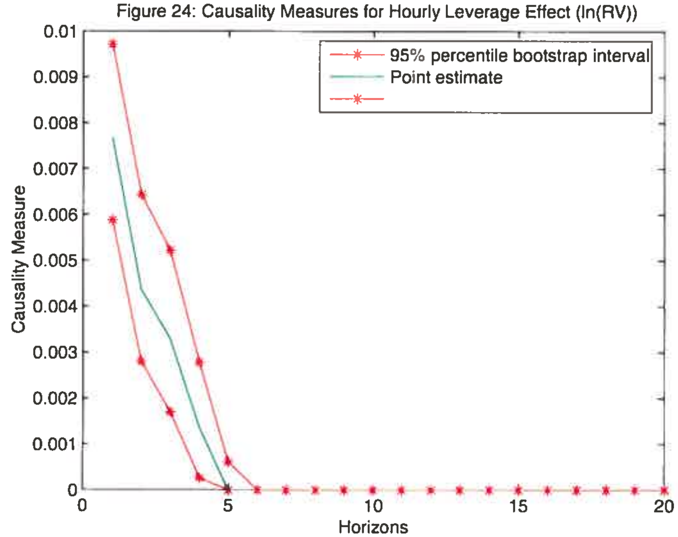


Figure 25: Causality Measures for Hourly Leverage Effect (ln(BV))

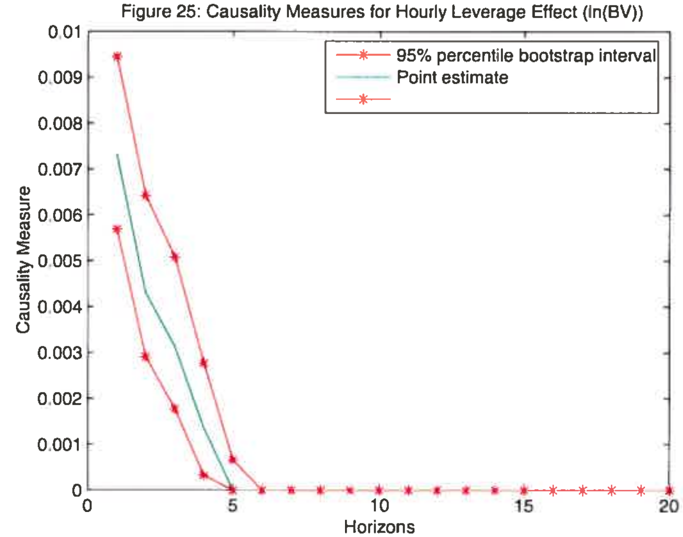


Figure 26: Measures of instantanous causality between daily return and relaized volatility

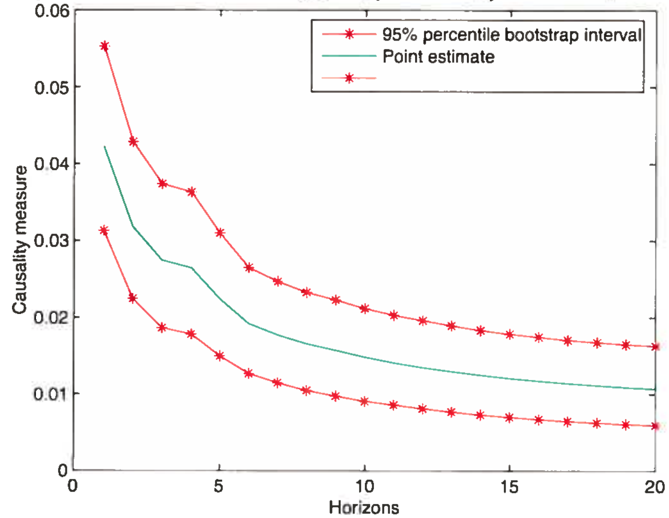


Figure 27: Measures of instantanous causality between daily return and Bipower variation

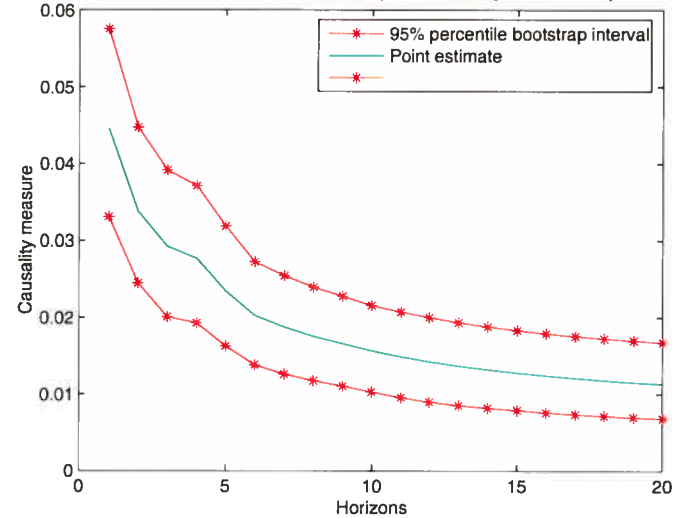


Figure 28: Measures of dependence between daily return and relaized volatility

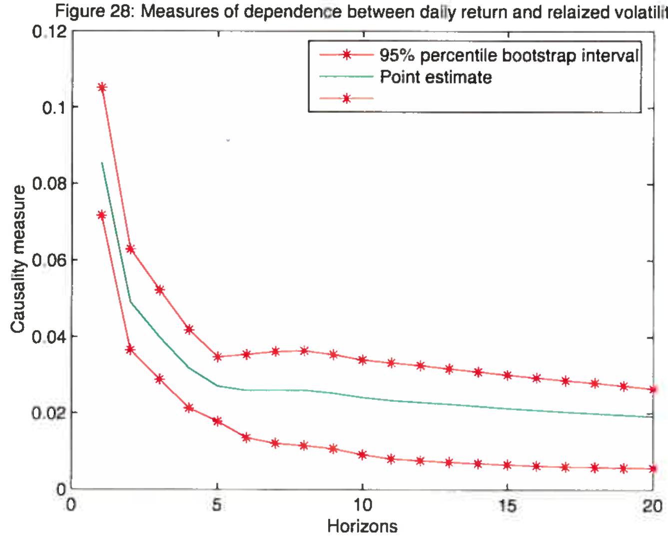


Figure 29: Measures of dependence between daily return and Bipower variation

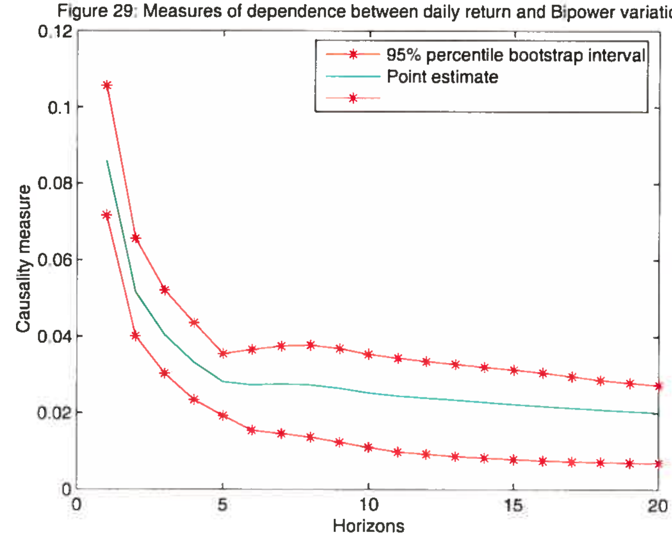


Figure 30. Comparison Between Daily Leverage and Feedback Effects (ln(RV))

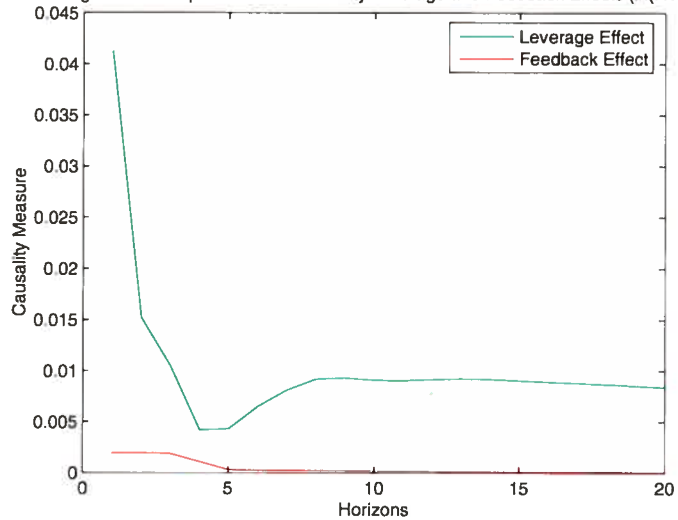


Figure 31: Comparison Between Daily Leverage and Feedback Effects (ln(BV))

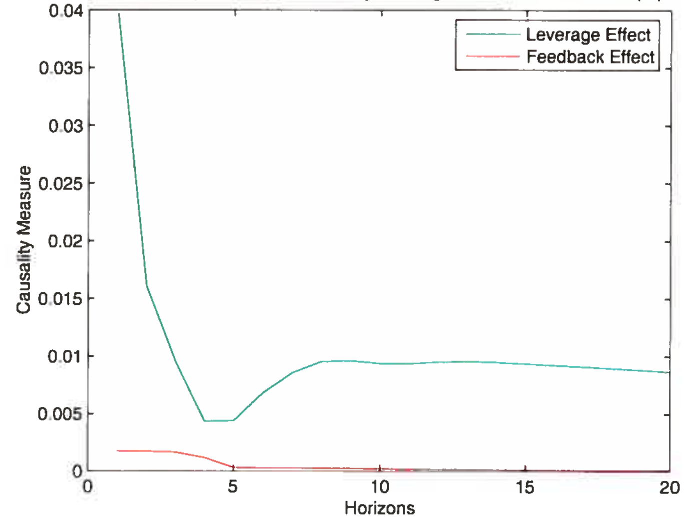


Figure 32: Hourly and Daily Leverage Effect ln(RV)

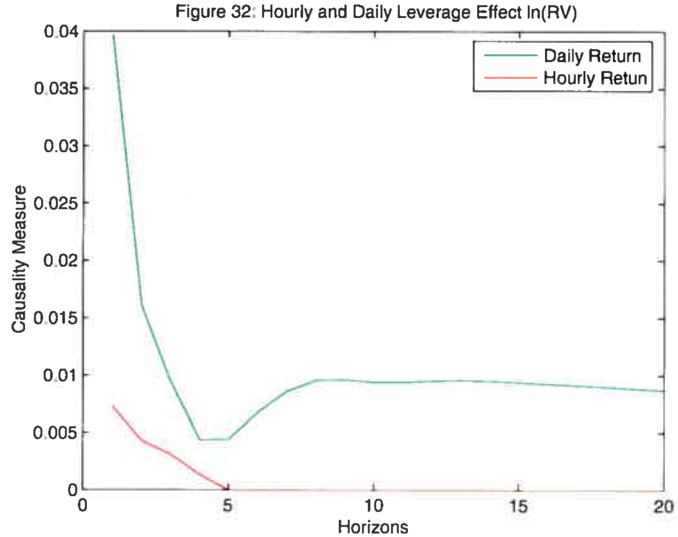


Figure 33: Hourly and Daily Leverage Effect ln(BV)

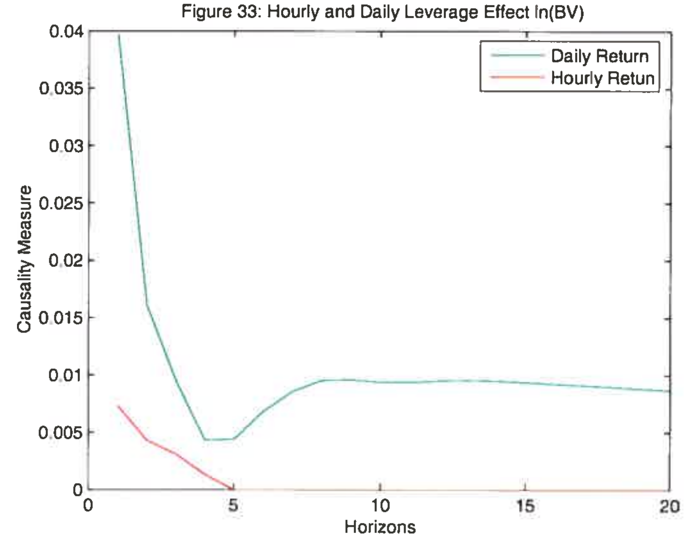


Figure 34: Impact of bad news on volatility (ln(RV) and m=15 jours)

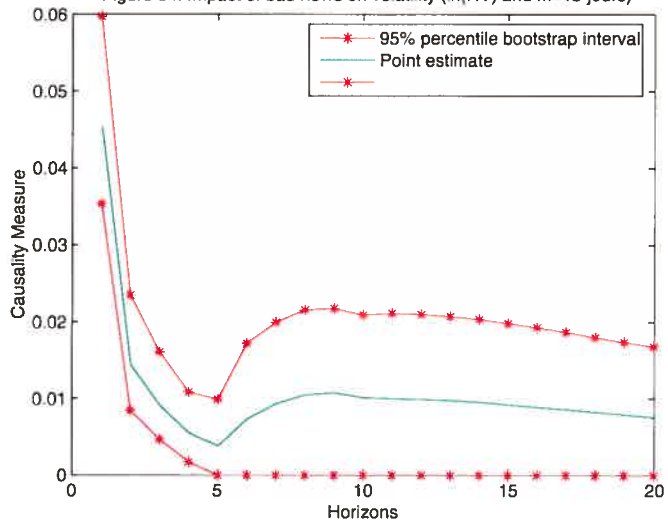


Figure 35: Impact of bad news on volatility (ln(BV) and m=15 jours)

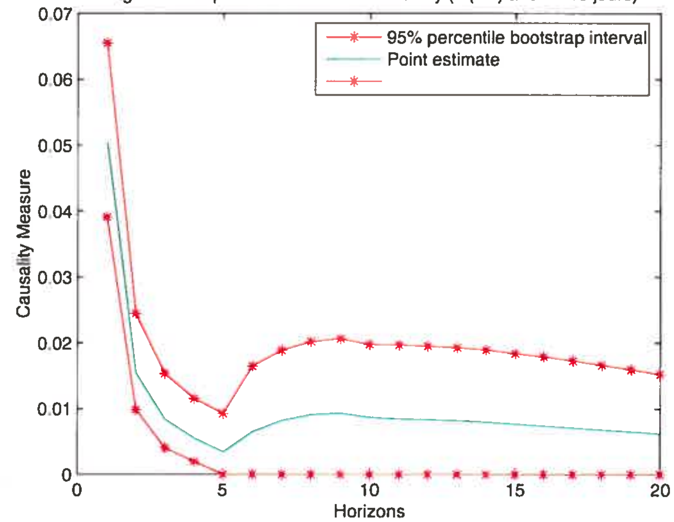


Figure 36: Impact of bad news on volatility (ln(RV) and m=30 jours)

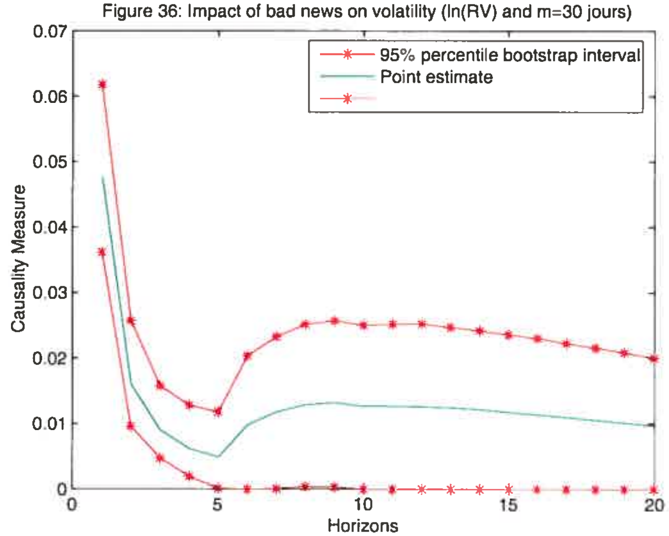


Figure 37: Impact of bad news on volatility (ln(BV) and m=30 jours)

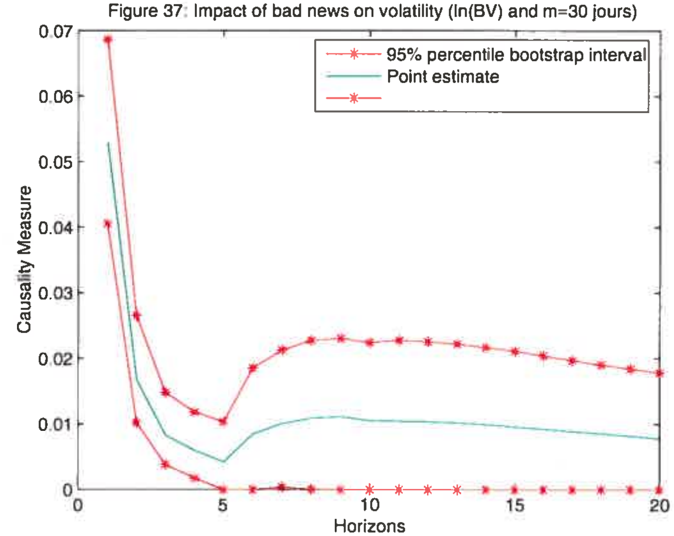


Figure 38: Impact of bad news on volatility (ln(RV) and m=90 jours)

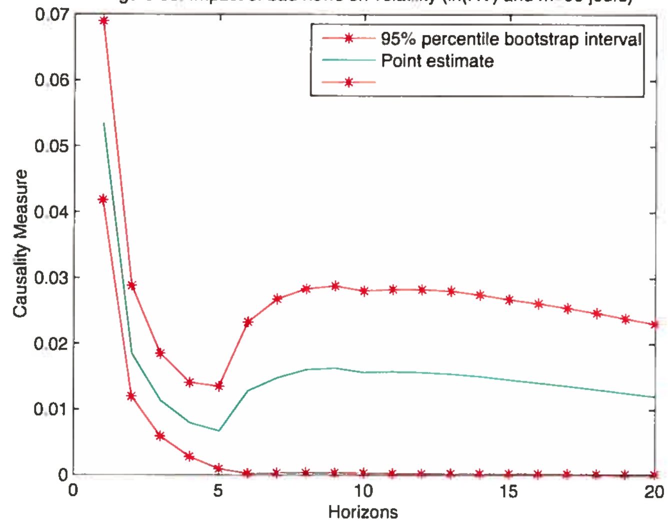


Figure 39: Impact of bad news on volatility (ln(BV) and m=90 jours)

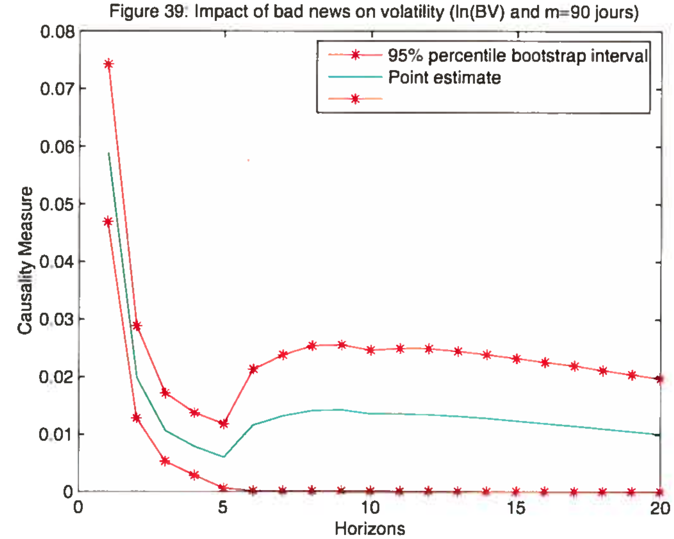


Figure 40: Impact of bad news on volatility (ln(RV) and m=120 jours)

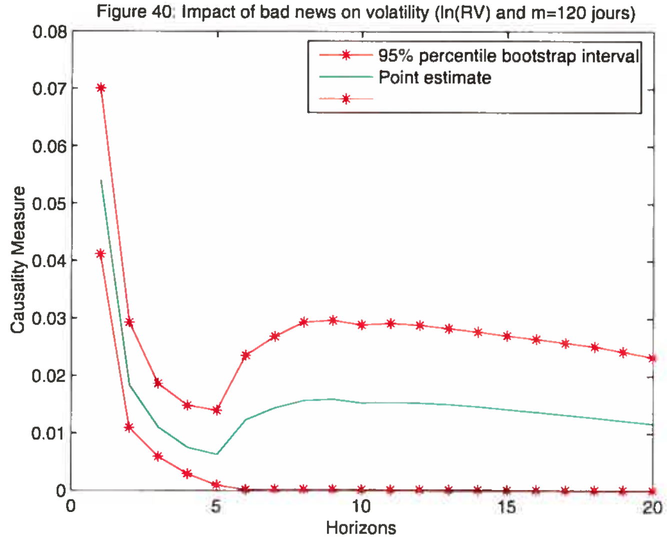


Figure 41: Impact of bad news on volatility (ln(BV) and m=120 jours)

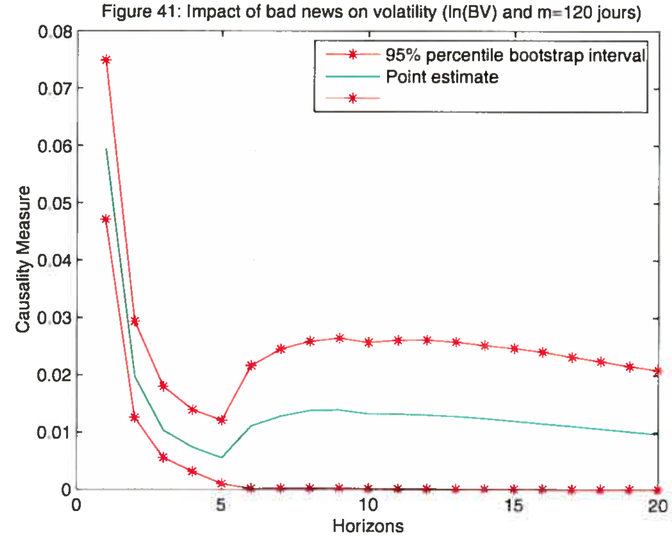


Figure 42: Impact of bad news on volatility (ln(RV) and m=240 jours)

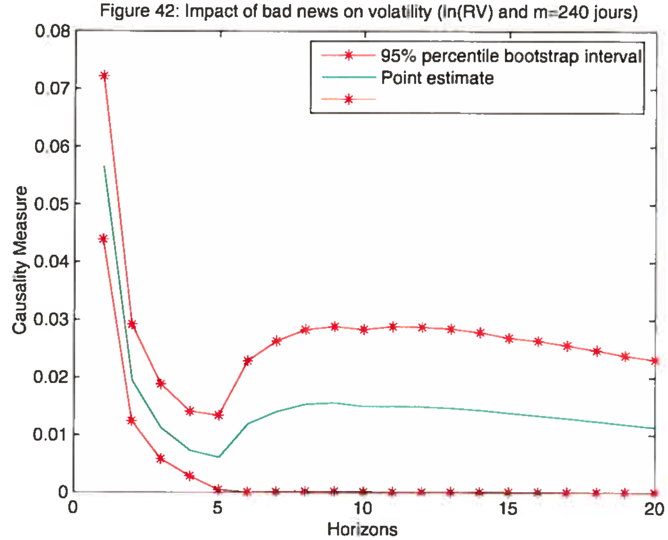


Figure 43: Impact of bad news on volatility (ln(BV) and m=240 jours)

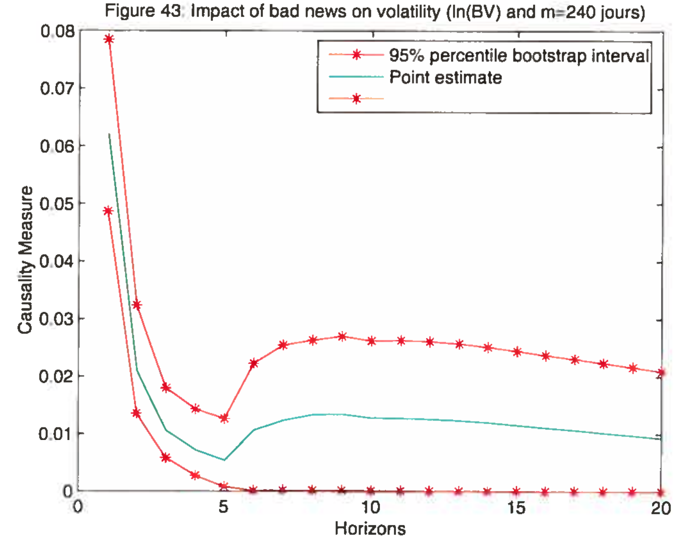


Figure 44: Impact of bad news on the volatility (ln(RV))

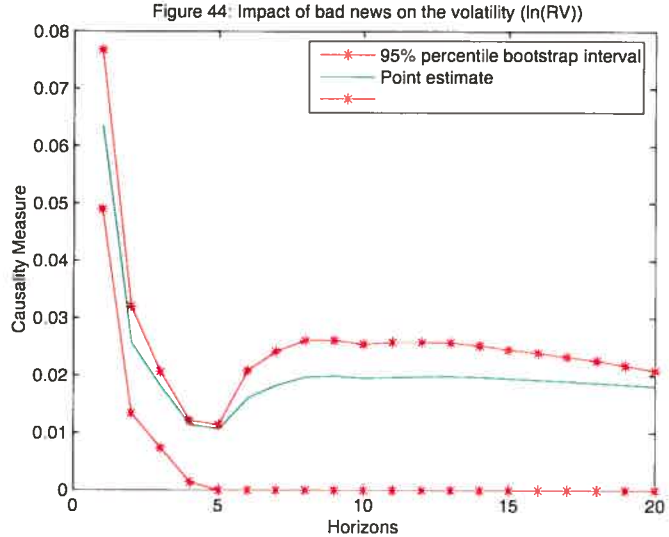


Figure 45: Impact of bad news on the volatility (ln(BV))

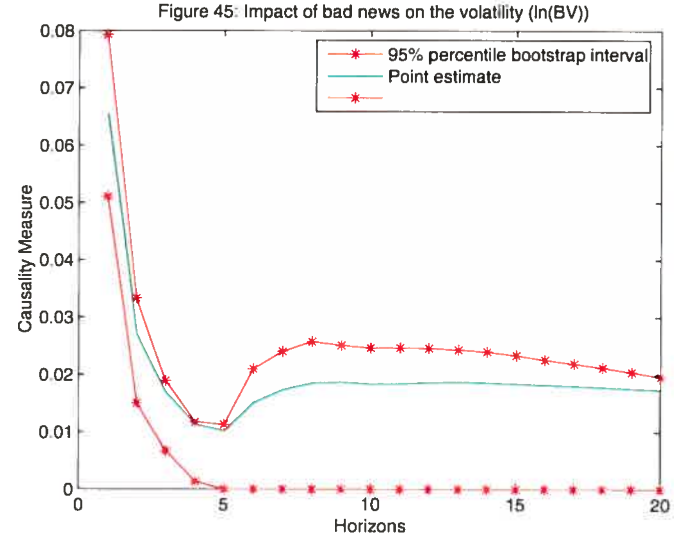


Figure 46: Comparing the impact of bad and good news on volatility (ln(RV))

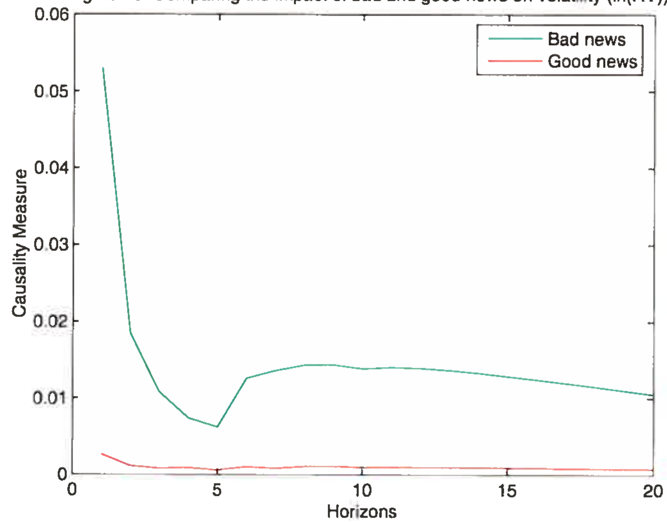


Figure 47: Comparing the impact of bad and good news on volatility (ln(BV))

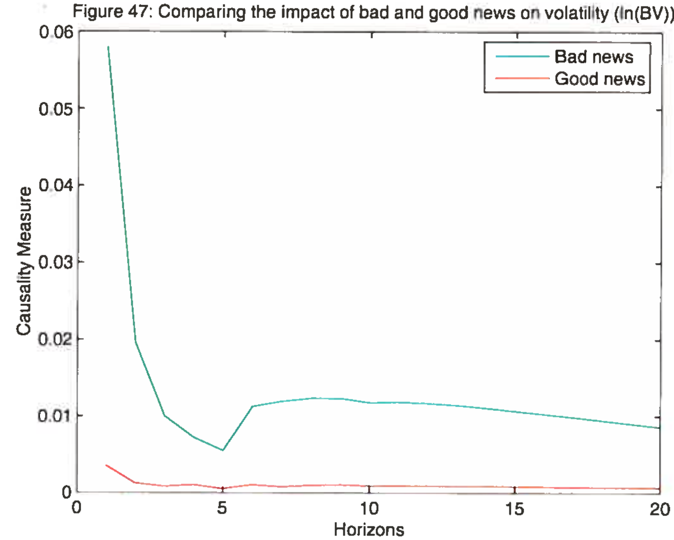


Figure 48: Difference between the impact of bad and good news on volatility (ln(BV))

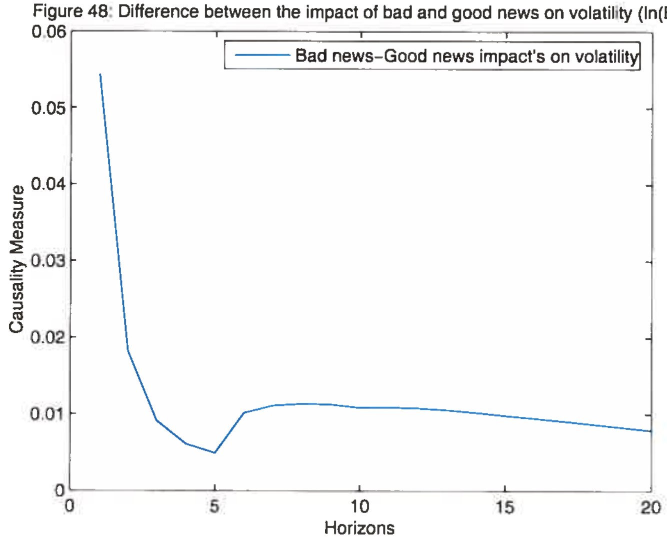


Figure 49: Difference between the impact of bad and good news on volatility (ln(BV))

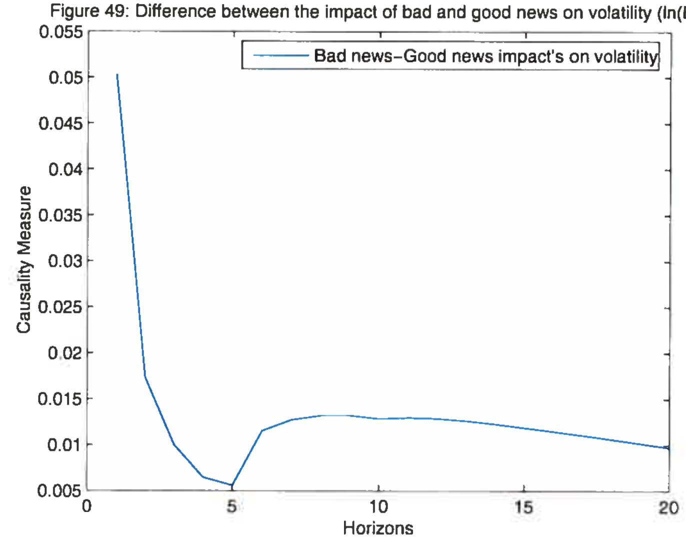


Figure 50: Temporal Aggregation and dependence between volatility and return (ln(BV))

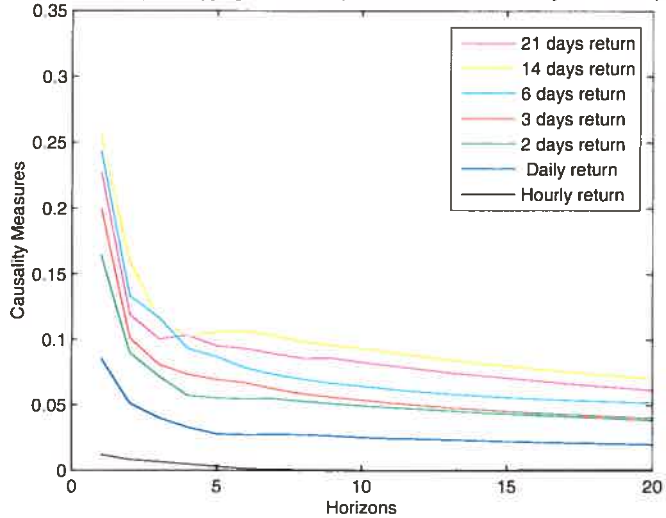
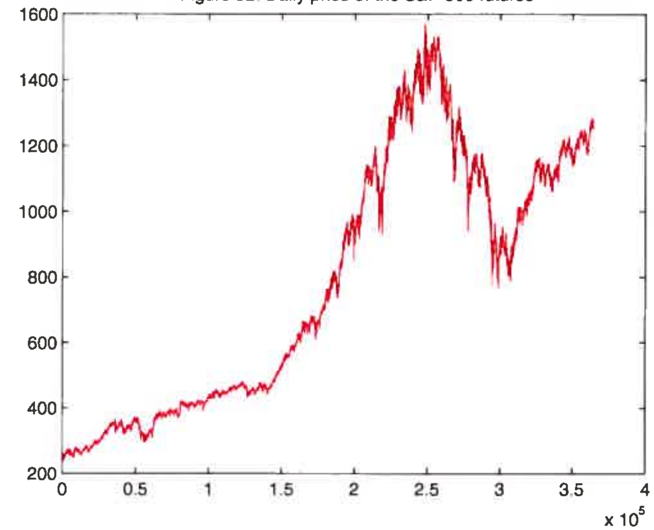


Figure 52: Daily price of the S&P 500 futures



Chapter 3

Risk measures and portfolio optimization under a regime switching model

3.1 Introduction

Since the seminal work of Hamilton (1989), Markov switching models have been increasingly used in financial time-series econometrics because of their ability to capture some key features, such as heavy tails, persistence, and nonlinear dynamics in asset returns. In this chapter, we exploit the superiority of these models to derive some financial risk measures, such as Value-at-Risk (VaR) and Expected Shortfall (ES), which take into account important stylized facts that we observe in equity markets. We also characterize the multi-horizon mean-variance efficient frontier of the linear portfolio and we compare the performance of the conditional and unconditional optimal portfolios.

VaR has become the most widely used technique to measure and control market risk. It is a quantile measure that quantifies risk for financial institutions and measures the worst expected loss over a given horizon (typically a day or a week) at a given statistical confidence level (typically 1%, 5% or 10%). Different methods exist to estimate VaR under different models of risk factors. Generally, there is a trade-off between the simplicity of the estimation method and the realism of the assumptions in the risk factor model: as we allow the latter to capture more stylized effects, the estimation method becomes more complex. Under the assumption that returns follow a conditional normal distribution, one can show that the VaR is given by a simple analytical formula [see RiskMetrics (1995)]. However, when we relax this assumption, the analytical calculation of the VaR becomes complicated and people tend to use computer-intensive simulation based methods. Based on the Markov switching model, this chapter proposes an analytical approximation of the VaR under more realistic assumptions than conditional normality.

The issue of VaR estimation under Markov switching regimes has been considered by Billio and Pelizzon (2000) and Guidolin and Timmermann (2005). Billio and Pelizzon (2000) use a switching volatility model to forecast the distribution of returns and to estimate the VaR of both single assets and linear portfolios. Comparing the calculated VaR values with the variance-covariance approach and $GARCH(1,1)$ models, they find that VaR values under switching regime models are preferable to the values under the other two methods. Guidolin and Timmermann (2005) examine the term structure of VaR under different econometric approaches, including multivariate regime switching, and find that bootstrap and regime switching models are best overall for VaR levels of 5% and 1%, respectively. To our knowledge, no analytical method has been proposed to estimate the VaR under Markov switching regimes. The present chapter uses the same approach as Cardenas et al. (1997), Rouvinez (1997), and Duffie and Pan (2001) to provide an analytical approximation to a multi-horizon conditional VaR under regime switching model. Using the Fourier inversion method, we first derive the probability distribution function for multi-horizon portfolio returns. Thereafter, we use an efficient numerical integration step, designed by Davies (1980), to approximate the infinite integral in the inversion formula and make estimation of the VaR feasible. Finally, we use the Hamilton filter to compute the conditional VaR.

Despite its popularity among managers and regulators, the VaR measure has been criticized because, in general, it lacks consistency and ignores losses beyond the VaR level. Furthermore, it is not subadditive, which means that it penalizes diversification instead of rewarding it. Consequently, researchers have proposed a new risk measure, called Expected Shortfall, which is the conditional expectation of a loss given that the loss is beyond the VaR level. Contrary to VaR, Expected Shortfall is consistent, takes the frequency and severity of financial losses into account, and is additive. To our knowledge, no analytical formula has been derived for the Expected Shortfall measure under Markov switching regimes. In this chapter we use the Fourier inversion method to derive a closed-form solution for the multi-horizon conditional Expected Shortfall measure.

Another objective of this chapter is to study portfolio optimization under Markov switching regimes. In the literature there are two ways of considering the problem of portfolio optimization: static and dynamic. In the Mean-Variance framework, the difference between these two approaches is related to how we calculate the first two moments of asset returns. In the static approach, the structure of the optimal portfolio is chosen once and for all at the beginning of the period. One critical drawback of this approach is that it assumes a constant mean and variance of returns. In the dynamic approach, the structure of the optimal portfolio is continuously adjusted using the available information set. One advantage of this approach is that it allows exploitation of the predictability of the first and second moments of asset returns and hedging changes in the investment opportunity set.

Several recent studies examine the economic implications of return predictability on investors' asset allocation decisions and find that investors react differently when returns are predictable.[1] In those studies we distinguish between two approaches. The first one, which evaluates the economic benefits via ex ante calibration, concludes that return predictability can improve investors' decisions [see Kandel and Stambaugh (1996), Balduzzi and Lynch (1999), Lynch (2001), Gomes (2002), and Campbell, Chan, and Viceira (2002)]. The second approach, which evaluates the ex post performance of return predictability, finds mixed results. Breen, Glosten, and Jagannathan (1989) and Pesaran and Timmermann (1995) find that return predictability yields significant economic gains out of sample, whereas Cooper, Gutierrez, and Marcum (2001) and Cooper and Gulen (2001) do not find any economic significance. In the Mean-Variance framework, Jacobsen (1999) and Marquering and Verbeek (2001) find that the economic gains of exploiting return predictability are significant, whereas Handa and Tiwari (2004) find that the economic significance of return predictability is questionable.[2]

Recently, Campbell and Viceira (2005) examined the implications of the predictability of the asset returns for multi-horizon asset allocation using standard vector autoregressive model with constant variance-covariance structure for shocks. They find the changes in investment opportunities can alter the risk-return trade-off of bonds, stocks, and cash across investment horizons, and that asset return predictability has important effects on the variance and correlation structure of returns on stocks, bonds and T-bills across

[1] Numerous empirical works have asked whether stock returns can be predicted or not: see Fama and Schwert (1977), Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988), Fama and French (1988, 1989), and Hodrick (1992), among others.

[2] See Han (2005) for more discussion.

investment horizons. In this chapter we extend the model of Campbell and Viceira (2005) by allowing for regime-switching in the mean and variance of returns. However, we do not consider variables such as price-earnings ratios, interest rate, or yield spreads, to predict future returns, as Campbell and Viceira (2005) did. We derive the conditional and unconditional first two moments of the multi-horizon portfolio return that we use to compare the performance of the dynamic and static optimal portfolios. Using daily observations on S&P 500 and TSE 300 indices, we first find that the conditional risk (variance and VaR) per period of the multi-horizon optimal portfolio's returns, when plotted as a function of the horizon h, may be increasing or decreasing at intermediate horizons, and converges to a constant- the unconditional risk-at long enough horizons. Second, the efficient frontiers of the multi-horizon optimal portfolios are time varying. Finally, at short-term and in 73.56% of the sample the conditional optimal portfolio performs better then the unconditional one.

The remainder of this chapter is organized as follows. In section 3.2, we introduce some notations and we derive the conditional and unconditional Laplace Transform of Markov chains. In section 3.3, we specify our model and we derive the probability distribution function of multi-horizon returns. We use this probability distribution function to approximate the multi-horizon portfolio's conditional VaR and derive a closed-form solution for the portfolio's conditional Expected Shortfall. In section 3.4, we characterize the multi-horizon mean-variance efficient frontier of the optimal portfolio under Markov switching regimes. A description of the data and the empirical results are given in section 3.5. We conclude in section 3.6. Technical proof are given in section 3.7.

3.2 Framework

In this section, we introduce some notations and we derive the conditional and unconditional Laplace Transform of simple and aggregated Markov chains. We assume that

$$\zeta_t = \begin{cases} (1,0,0,....,0)^\top \text{ when } s_t = 1 \\ (0,1,0,....,0)^\top \text{ when } s_t = 2 \\ . \\ . \\ . \\ (0,0,0,....,1)^\top \text{ when } s_t = N \end{cases}$$

where s_t is a stationary and homogenous Markov chain. It is well known that (see, e.g., Hamilton (1994), page 679)

$$E[\zeta_{t+h} \mid J_t] = P^h \zeta_t, h \geq 1, \tag{3.1}$$

where J_t is an information set and

$$P = [p_{ij}]_{1\leq i,j\leq N}\ ,\ p_{ij} = \mathsf{P}(s_{t+1} = j \mid s_t = i). \tag{3.2}$$

We assume that the Markov chain is stationary with an ergodic distribution Π, $\Pi\in \mathbb{R}^N$, i.e.

$$E[\zeta_t] = \mathbf{\Pi}. \tag{3.3}$$

Observe that

$$P^h \mathbf{\Pi} = \mathbf{\Pi},\ \forall h. \tag{3.4}$$

In what follows, we adopt the notations:

$$A(u) = Diag\left(\exp(u_1), \exp(u_2), ..., \exp(u_N)\right) P, \ \forall u \in \mathbb{R}^N, \tag{3.5}$$

$$P^h = [P_{ij}(h)]_{1 \leq i,j \leq N}.$$

The conditional and unconditional Laplace Transform of simple and aggregated Markov chains are given by the following propositions.

Proposition 1 ***(Conditional Laplace Transform of Markov Chains)*** $\forall u \in \mathbb{R}^N, \forall h \geq 1$, *we have*

$$E[\exp(u^\top \zeta_{t+h}) \mid J_t] = e^\top A(u) P^{h-1} \zeta_t,$$

$$E[\exp(u^\top \zeta_{t+1}) \zeta_{t+1} \mid J_t] = A(u) \zeta_t.$$

Proposition 2 ***(Joint Laplace Transform of the Markov Chain)*** $\forall h \geq 1, \forall u_i \in \mathbb{R}^N$, *for* $i = 1, ..., h$, *we have*

$$E[\exp\left(\sum_{i=1}^{h} u_i^\top \zeta_{t+i}\right) \mid J_t] = e^\top \prod_{\mathbf{i=1}}^{\mathbf{h}} A(u_{h+1-i}) \zeta_t, \tag{3.6}$$

$$E[\exp\left(\sum_{i=1}^{h} u_i^\top \zeta_{t+i}\right)] = e^\top \prod_{\mathbf{i=1}}^{\mathbf{h}} A(u_{h+1-i}) \Pi. \tag{3.7}$$

where e *denotes the* $N \times 1$ *vector whose all components equal one.*

3.3 VaR and Expected Shortfall under Markov Switching regimes

There are n risky assets in the economy, the prices of which are given by $P_t = (P_{1t}, P_{2t}, ..., P_{nt})^\top$. We denote by $r_t = (r_{1t}, r_{2t}, ..., r_{nt})^\top$, where $r_{it} = \ln(P_{it}) - \ln(P_{i(t-1)})$ for $i = 1, ..., n$, the vector of assets returns. We define the information sets as:

$$J_t = \sigma(r_\tau, \zeta_\tau, \tau \leq t) = \sigma(r_\tau, s_\tau, \tau \leq t),$$

$$I_t = \sigma(r_\tau, \tau \leq t).$$

We assume that r_t follows a multivariate Markov switching model,

$$r_{t+1} = \mu\zeta_t + \Sigma(\zeta_t)\varepsilon_{t+1}, \quad \varepsilon_{t+1} \; i.i.d. \sim \mathcal{N}(0, I_n), \tag{3.8}$$

$$E\left[\Sigma(\zeta_t)\varepsilon_{t+1}\varepsilon_{t+1}^{\top}\Sigma(\zeta_t)^{\top} \mid J_t\right] = \Sigma(\zeta_t)I_n\Sigma(\zeta_t)^{\top} = \Omega(\zeta_t),$$

where I_n is an $n \times n$ identity matrix and

$$\mu = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1N} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2N} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nN} \end{pmatrix}, \quad \Omega(\zeta_t) = \begin{pmatrix} \omega_{11}^{\top}\zeta_t & \omega_{12}^{\top}\zeta_t & \cdots & \omega_{1n}^{\top}\zeta_t \\ \omega_{21}^{\top}\zeta_t & \omega_{22}^{\top}\zeta_t & \cdots & \omega_{2n}^{\top}\zeta_t \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \omega_{n1}^{\top}\zeta_t & \omega_{n2}^{\top}\zeta_t & \cdots & \omega_{nn}^{\top}\zeta_t \end{pmatrix}.$$

μ_{ij}, for $i = 1, ..., n$ and $j = 1, ..., N$, is the mean return of an asset i at state j and ω_{il}, for $i, l = 1, ..., n$, is a vector of covariances between assets i and l at the N states. The processes $\{s_t\}$ and $\{\varepsilon_t\}$ are assumed jointly independent.

3.3.1 One-period-ahead VaR and Expected Shortfall

To compute the VaR of linear portfolio, we proceed in three steps. First, we calculate the characteristic function of the portfolio's return. Second, we follow Gil-Pelaez (1951) and use the Fourier inversion method to compute the probability distribution of the portfolio's return. Third, we compute the VaR by inverting the probability distribution function and using an efficient numerical integration step designed by Davies (1980). We also use the Fourier inversion method to derive a closed-form solution of the Expected Shortfall measure. Let us consider a linear portfolio of n assets, the return of which at time $t+1$ is given by:

$$r_{p,t+1} = \sum_{i=1}^{n} \alpha_i r_{it+1} = W^\top r_{t+1}, \tag{3.9}$$

where $W = (\alpha_1, \alpha_2, ..., \alpha_n)^\top$ is a vector representing the weight attributed to each asset in portfolio. At the horizon one, the conditional characteristic function of $r_{p,t+1}$ is given by the following proposition.

Proposition 3 (*Conditional Characteristic Function*) *$\forall\ u \in \mathbb{R}$, we have*

$$E[\exp(iur_{p,t+1}) \mid J_t] = \exp\left(\left(iu\mu^\top W - \frac{u^2}{2} \sum_{1 \le l_1, l_2 \le n} \alpha_{l_1} \alpha_{l_2} \omega_{l_1 l_2}\right)^\top \zeta_t\right). \tag{3.10}$$

where i is a complex variable such that $i = \sqrt{-1}$.

The function (3.10) depends on the state variable ζ_t which is not observable. In practice, we need to filter this function using an observable information set. Using the law of iterated expectations, we get

$$E[\exp(iur_{p,t+1}) \mid I_t] = E[\exp\left(\left(iu\mu^\top W - \tfrac{u^2}{2} \textstyle\sum_{1 \le l_1, l_2 \le n} \alpha_{l_1} \alpha_{l_2} \omega_{l_1 l_2}\right)^\top \zeta_t\right) \mid I_t]$$

$$= \textstyle\sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \exp\left(iuW^\top \mu_j - \tfrac{u^2}{2}(W^\top \Omega_j W)\right),$$

where I_t is the observable information set, μ_j is the $n \times 1$ mean return vector at state j, and Ω_j is the $n \times n$ variance-covariance matrix of the n assets' returns at state j. According to Gil-Pelaez (1951), the conditional distribution function of $r_{p,t+1}$ evaluated at $\bar{r}$, for $\bar{r} \in \mathbb{R}$, is given by:

$$\mathsf{P}_t(r_{p,t+1} < \bar{r}) = \frac{1}{2} - \frac{1}{\pi} \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \int_0^{+\infty} \frac{I_j(u)}{u} du, \tag{3.11}$$

where[3]

$$I_j(u) = \mathrm{Im}\left\{\exp\left(iuW^\top\mu_j - \frac{u^2}{2}(W^\top\Omega_j W)\right)\exp(-iu\bar{r})\right\}.$$

$\mathrm{Im}(z)$ denotes the imaginary part of a complex number z. We have,

$$I_j(u) = \exp\left(-u^2 W^\top\Omega_j W/2\right)\sin\left(u(W^\top\mu_j - \bar{r})\right).$$

In what follows we assume that the VaR is a positive quantity:

$$\mathsf{P}_t(r_{p,t+1} < -VaR) = \frac{1}{2} - \frac{1}{\pi}\sum_{j=1}^{N}\mathsf{P}(s_t = j \mid I_t)\int_0^\infty \frac{I_j(u)}{u}du, \tag{3.12}$$

where

$$I_j(u) = \exp\left(-u^2 W^\top\Omega_j W/2\right)\sin\left(u(W^\top\mu_j + VaR)\right).$$

The VaR is a quantile measure and it can be computed by inverting the distribution function (3.12). However, inverting equation (3.12) analytically is not feasible and a numerical approach is required.

Proposition 4 (*Conditional* VaR) *The one-period-ahead portfolio's conditional-VaR with coverage probability α, denoted $VaR_t^\alpha(r_{p,t+1})$, is the solution of the following equation*

$$\sum_{j=1}^{N}\mathsf{P}(s_t = j \mid I_t)\int_0^\infty \frac{I_j(u)}{u}du - (\frac{1}{2} - \alpha)\pi = 0 \tag{3.13}$$

where, for $j = 1, \ldots, N$,

$$I_j(u) = \exp\left(-\frac{u^2}{2}(W^\top\Omega_j W)\right)\sin\left(u(W^\top\mu_j + VaR_t^\alpha(r_{p,t+1}))\right).$$

Corollary 3 (*Unconditional* VaR) *The one-period-ahead portfolio's unconditional-VaR*

[3]The subscript t in the probability distribution function (3.11) is to indicate that we condition on the information set I_t.

with coverage probability α, *denoted* $VaR^{\alpha}(r_{p,t+1})$, *is the solution of the following equation*

$$\sum_{j=1}^{N} \pi_j \int_0^{\infty} \frac{I_j(u)}{u} du - (\frac{1}{2} - \alpha)\pi = 0,$$

where π_j, *for* $j = 1, ..., N$, *are the ergodic or steady state probabilities.*

Corollary **3** can be deduced from Proposition **4** using the law of iterated expectations. The conditional VaR can be approximated by solving the function

$$f(VaR^{\alpha}) = \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \int_0^{\infty} \frac{I_j(u)}{u} du - (\frac{1}{2} - \alpha)\pi = 0. \tag{3.14}$$

The function $f(VaR^{\alpha})$ can be written in the following form:

$$f(VaR^{\alpha}) = -\pi[\mathsf{P}_t(r_{p,t+1} < -VaR^{\alpha}) - \alpha]. \tag{3.15}$$

Using the properties of the probability distribution function (monotonically increasing, $\lim_{x\to-\infty} \mathsf{P}_t(r_{p,t+1} < x) = 0$, and $\lim_{x\to+\infty} \mathsf{P}_t(r_{p,t+1} < x) = 1$) one can show that (3.15) has a unique solution [see proof in appendix 2]. Another way to approximate the conditional VaR is to consider the following optimization problem:

$$\widehat{VaR_t^{\alpha}(r_{p,t+1})} = \underset{VaR_t^{\alpha}}{Arg\min} \left[(\frac{1}{2} - \alpha)\pi - \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \int_0^{\infty} \frac{I_j(u)}{u} du \right]^2, \tag{3.16}$$

where

$$I_j(u) = \exp\left(-u^2(W^{\top}\Omega_j W)/2\right) \sin\left(u(W^{\top}\mu_j + VaR_t^{\alpha})\right).$$

The following is an algorithm that one can follow to compute the portfolio's conditional-VaR:

1. Estimate the vector of the unknown parameters

$$\theta = \left(vec(\mu)^{\top}, vech(\Omega_1)^{\top}, ..., vech(\Omega_N)^{\top}, vec(P)^{\top}\right)^{\top}$$

using the maximum-likelihood method [see Hamilton (1994, pages 690–696)],

2. Estimate the conditional probability of regimes,

$$\Pi_{s_t} = \hat{\xi}_{t+1|t} = (\mathsf{P}(s_t = 1 \mid I_t), ..., \mathsf{P}(s_t = N \mid I_t))^\top ,$$

by iterating on the following pair of equations [see Hamilton (1994)]:

$$\hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t|t-1} \odot \eta_t)}{e^\perp(\hat{\xi}_{t|t-1} \odot \eta_t)}, \tag{3.17}$$

$$\hat{\xi}_{t+1|t} = P\hat{\xi}_{t|t}. \tag{3.18}$$

where, for $t = 1, ..., T$,

$$\hat{\xi}_{t|t} = (\mathsf{P}(s_{t-1} = 1 \mid I_t), ..., \mathsf{P}(s_{t-1} = N \mid I_t))^\top ,$$

$$\eta_t = \begin{bmatrix} \frac{1}{\sqrt{2\pi(W^\top \Omega_1 W)}} \exp\left\{\frac{-(r_{p,t} - W^\top \mu_1)^2}{(W^\top \Omega_1 W)}\right\} \\ \frac{1}{\sqrt{2\pi(W^\top \Omega_2 W)}} \exp\left\{\frac{-(r_{p,t} - W^\top \mu_2)^2}{(W^\top \Omega_2 W)}\right\} \\ \cdots \\ \frac{1}{\sqrt{2\pi(W^\top \Omega_N W)}} \exp\left\{\frac{-(r_{p,t} - W^\top \mu_N)^2}{(W^\top \Omega_N W)}\right\} \end{bmatrix},$$

the symbol $\odot$ denotes element-by-element multiplication. Given a starting value $\hat{\xi}_{1|0}$ and the estimator $\hat{\theta}^{MV}$ of the vector θ, one can iterate on (3.17) and (3.18) to compute the values of $\hat{\xi}_{t|t}$ and $\hat{\xi}_{t+1|t}$ for each date t in the sample. Hamilton (1994, pages 693–694) suggests several options for choosing the starting value $\hat{\xi}_{1|0}$. One approach is to set $\hat{\xi}_{1|0}$ equal to the vector of unconditional probabilities Π. Another option is to set $\hat{\xi}_{1|0} = \boldsymbol{\rho}$, where $\boldsymbol{\rho}$ is a fixed $N \times 1$ vector of nonnegative constants summing to unity, such as $\boldsymbol{\rho} = N^{-1}e$. Alternatively, $\boldsymbol{\rho}$ can be estimated by maximum likelihood, along with θ, subject to the constraint that $e^\perp\boldsymbol{\rho} = 1$ and $\boldsymbol{\rho}_j \geq 0$ for $j = 1, 2, ..., N$.

3. Given $\hat{\theta}^{MV}$ and Π_{s_t}, the portfolio's conditional-VaR with coverage probability α is the solution to the following optimization problem:

$$\widehat{VaR_t^\alpha(r_{p,t+1})} = \underset{VaR_t^\alpha}{Arg\ \min} \left[(\frac{1}{2} - \alpha)\pi - \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \int_0^\infty \frac{I_j(u)}{u} du \right]^2 \qquad (3.19)$$

where

$$I_j(u) = \exp\left(-u^2(W^\top \hat{\Omega}_j^{MV} W)/2\right) \sin\left(u(W^\top \hat{\mu}_j^{MV} + VaR_t^\alpha)\right).$$

In practice, an exact solution of equation (??) is not feasible, since the integral $\int_0^\infty \frac{I_j(u)}{u} du$ is difficult to evaluate. The latter can be approximated using results by Imhof (1961), Bohmann (1961, 1970, 1972), and Davies (1973), who propose a numerical approximation of the distribution function using the characteristic function. The proposed approximation introduces two types of errors: discretization and truncation errors. Davies (1973), proposes a criterion for controlling discretization error and Davies (1980) proposes three different bounds for controlling truncation error. Furthermore, Shephard (1991a,b) provides rules for the numerical inversion of a multivariate characteristic function to compute the distribution function. These rules represent a multivariate generalization of the Imhof (1961) and Davies (1973, 1980) results.

The VaR measure has been criticized for several reasons; it lacks consistency, ignores losses beyond the VaR level, and it is not subadditive, which means that it penalizes diversification instead of rewarding it. Consequently, researchers have proposed a new risk measure, called the Expected Shortfall, which is the conditional expectation of loss given that the loss is beyond the VaR level. Unlike the VaR, Expected Shortfall is consistent, takes the frequency and severity of financial losses into account, and is additive. Given its importance for evaluating financial market risk, the following propositions give a closed-form solution for the portfolio's Expected Shortfall measure.

Proposition 5 (*Conditional Expected Shortfall*) *The one-period-ahead portfolio's* *conditional-Expected Shortfall with coverage probability* α, *denoted* $ES_t^\alpha(r_{p,t+1})$, *is given by:*

$$ES_t^\alpha(r_{p,t+1}) = \frac{1}{\alpha\sqrt{2\pi}} e^\top R(u)\Pi_{s_t},$$

where

$$R(u) = Diag\left[\exp\left(-\frac{1}{2}\frac{(W^\top \mu_1 + VaR_t(r_{p,t+1}))^2}{(W^\top \Omega_1 W)}\right), ..., \exp(-\frac{1}{2}\frac{(W^\top \mu_N + VaR_t(r_{p,t+1}))^2}{(W^\top \Omega_N W)}\right)\Bigg].$$

Corollary 4 (***Unconditional Expected Shortfall***) *The one-period-ahead portfolio's unconditional-Expected Shortfall with coverage probability* α, *denoted* $ES^\alpha(r_{p,t+1})$, *is given by*

$$ES^\alpha(r_{p,t+1}) = \frac{1}{\alpha\sqrt{2\pi}} e^\top R(u)\Pi,$$

where

$$R(u) = Diag\left[\exp\left(-\frac{1}{2}\frac{(W^\top \mu_1 + VaR(r_{p,t+1}))^2}{(W^\top \Omega_1 W)}\right), ..., \exp(-\frac{1}{2}\frac{(W^\top \mu_N + VaR(r_{p,t+1}))^2}{(W^\top \Omega_N W)}\right)\Bigg].$$

Corollary **4** can be deduced from Proposition **5** using the law of iterated expectations.

3.3.2 Multi-Horizon VaR and Expected Shortfall

We denote by $r_{t:t+h} = \sum_{k=1}^{h} r_{t+k}$ the multi-horizon aggregated return, where r_{t+k} follows a multivariate Markov switching model (3.8). To compute the multi-horizon VaR and Expected Shortfall of linear portfolio, we follow the same steps as in subsection (3.3.1). Based on Propositions **1** and **2**, the characteristic functions of the h-period-ahead portfolio's return and aggregated portfolio's return are given by the following proposition.

Proposition 6 (***Multi-horizon Conditional Characteristic Function***) $\forall\, u \in \mathbb{R}$ *and*

$h \geq 2$, *we have*

$$E[\exp(iur_{p,t+h}) \mid J_t] = e^{\top} A\left(iu\mu^{\top}W - \frac{u^2}{2}\sum_{1\leq l_1,l_2\leq n}\alpha_{l_1}\alpha_{l_2}\omega_{l_1l_2}\right)P^{h-2}\zeta_t. \quad (3.20)$$

$$E[\exp(iur_{p,t:t+h}) \mid J_t] = e^{\top}\left(A\left(iu\mu^{\top}W - \frac{u^2}{2}\sum_{1\leq l_1,l_2\leq n}\alpha_{l_1}\alpha_{l_2}\omega_{l_1l_2}\right)\right)^{h-1} \quad (3.21)$$

$$\times \exp\left(\left(iu\mu^{\top}W - \frac{u^2}{2}\sum_{1\leq l_1,l_2\leq n}\alpha_{l_1}\alpha_{l_2}\omega_{l_1l_2}\right)^{\top}\zeta_t\right)\zeta_t$$

where

$$\mathrm{A}\left(iu\mu^{\top}\mathrm{W}\text{-}\frac{u^2}{2}\sum_{1\leq l_1,l_2\leq n}\alpha_{l_1}\alpha_{l_2}\omega_{l_1l_2}\right) = Diag\left(\exp\left(a_1\right)\ldots..\exp\left(a_N\right)\right),$$

and, for $j = 1, ..., N$,

$$a_j = iuW^{\top}\mu_j - \frac{u^2}{2}W^{\top}\Omega_j W,$$

e denotes the $N \times 1$ *vector whose components are all equal to one.*

The functions (3.20)-(3.21) depend on the state variable ζ_t. In practice, the current state variable ζ_t is not observable and one needs to use the observable information set I_t to filter these functions. For the h-period-ahead portfolio's return, the law of iterated expectations yields

$$E[\exp(iur_{p,t+h}) \mid I_t] = e^{\top} A\left(iu\mu^{\top}W - \frac{u^2}{2}\sum_{1\leq l_1,l_2\leq n}\alpha_{l_1}\alpha_{l_2}\omega_{l_1l_2}\right)P^{h-2}\Pi_{s_t} \quad (3.22)$$

where

$$\Pi_{s_t} = \left(\mathrm{P}[s_t = 1 \mid I_t], ..., \mathrm{P}[s_t = N \mid I_t]\right)^{\top}.$$

An estimate of Π_{s_t} can be obtained by iterating on (3.17) and (3.18). Equation (3.22) is a complex function and it can be written as follows:

$$E[\exp(iur_{p,t+h}) \mid I_t] = e^\top[A_1(u) + iA_2(u)]P^{h-1}\Pi_{s_t}.$$

where

$$A_1(u)=Diag\left(\exp(\tfrac{-u^2}{2}W^\top\Omega_1 W)\cos(uW^\top\mu_1),\dots,\exp(\tfrac{-u^2}{2}W^\top\Omega_N W)\cos(uW^\top\mu_N)\right),$$

$$A_2(u)=\text{Diag}\left(\exp(\tfrac{-u^2}{2}W^\top\Omega_1 W)\sin(uW^\top\mu_1),\dots,\exp(\tfrac{-u^2}{2}W^\top\Omega_N W)\sin(uW^\top\mu_N)\right).$$

Similarly, the characteristic function of the h-period-ahead aggregated portfolio return is given by:

$$E[\exp(iur_{p,t:t+h}) \mid I_t] = e^\top\left(A\left(iu\mu^\top W - \frac{u^2}{2}\sum_{1\le l_1,l_2\le n}\alpha_{l_1}w\alpha_{l_2}\omega_{l_1l_2}\right)\right)^{h-1} D(u)\Pi_{s_t}. \tag{3.23}$$

where

$$D(u)=Diag\left(\exp\left(iuW\mu_1^\top-\frac{u^2}{2}W^\top\Omega_1 W\right),\dots,\exp\left(iuW\mu_N^\top-\frac{u^2}{2}W^\top\Omega_N W\right)\right).$$

which can be written as follows:

$$E[\exp(iur_{p,t:t+h}) \mid I_t] = e^\top[D_1(u) + iD_2(u)]\Pi_{s_t}.$$

where

$$D_1(u) = \text{Re}\left(\left(A\left(iu\mu^\top W - \tfrac{u^2}{2}\textstyle\sum_{1\le l_1,l_2\le n}\alpha_{l_1}\alpha_{l_2}\omega_{l_1l_2}\right)\right)^{h-1} D(u)\right).$$

$$D_2(u) = \text{Im}\left(\left(A\left(iu\mu^\top W - \tfrac{u^2}{2}\textstyle\sum_{1\le l_1,l_2\le n}\alpha_{l_1}\alpha_{l_2}\omega_{l_1l_2}\right)\right)^{h-1} D(u)\right).$$

Re(M) and Im(M) denote the real and imaginary parts of a complex matrix M, respectively. According to Gil-Pelaez (1951), the conditional distribution function of $r_{p,t+h}$, evaluated at $\bar{r}_p$ for $\bar{r}_p \in \mathbb{R}$, is given by:

$$\mathsf{P}_t(r_{p,t+h} < \bar{r}_p) = \frac{1}{2} - \frac{1}{\pi} e^\top \int_0^\infty \frac{\bar{A}_2(u)}{u} du \; P^{h-1}\Pi_{s_t}$$

where

$$\bar{A}_2(u) = Diag\left(\exp(\frac{-u^2}{2} W^\top \Omega_1 W)\sin(u(W^\top \mu_1 - \bar{r}_p)), \ldots, \exp(\frac{-u^2}{2} W^\top \Omega_N W)\sin(u(W^\top \mu_N - \bar{r}_p))\right).$$

Similarly, the conditional distribution function of $r_{p,t:t+h}$, evaluated at r_p for $r_p \in \mathbb{R}$, is given by:

$$\mathsf{P}_t(r_{p,t:t+h} < r_p) = \frac{1}{2} - \frac{1}{\pi} e^\top \int_0^\infty \frac{\bar{D}_2(u)}{u} du \; \Pi_{s_t},$$

where

$$\bar{D}_2(u) = \mathrm{Im}\left\{ \exp(-iur_p) A \left(iu\mu^\top W - \frac{u^2}{2} \sum_{1 \leq l_1, l_2 \leq n} \alpha_{l_1} \alpha_{l_2} \omega_{l_1 l_2} \right)^{h-1} D(u) \right\}.$$

It is not easy to have an explicit formula for the matrix $\bar{D}_2(u)$ as in the case of $\bar{A}_2(u)$. However, for a given finite horizon h, one can easily calculate the expression of $\bar{D}_2(u)$. Another way of calculating $\bar{D}_2(u)$ is to start by calculating $E[\exp(iur_{p,t+h}) \mid I_t]$ in term of sums and then to separate the imaginary and real parts of $E[\exp(iur_{p,t+h}) \mid I_t]$.

Proposition 7 (*Multi-Horizon Conditional VaR*) *The h-period-ahead portfolio's conditional-VaR with coverage probability α, denoted $VaR_t^\alpha(r_{p,t+h})$, is the solution of the following equation:*

$$e^\top \int_0^\infty \frac{\bar{A}_2(u)}{u} du \; P^{h-1}\Pi_{s_t} - (\alpha - \frac{1}{2})\pi = 0.$$

Similarly, the h-period-ahead aggregated portfolio's conditional-VaR with coverage probability α, denoted $VaR_t^\alpha(r_{p,t:t+h})$, is the solution of the following equation:

$$e^\top \int_0^\infty \frac{\bar{D}_2(u)}{u} du \; \Pi_{s_t} - (\alpha - \frac{1}{2})\pi = 0.$$

Corollary 5 (Multi-Horizon Unconditional VaR) *The h-period-ahead aggregated portfolio's unconditional-VaR with coverage probability α, denoted $VaR^{\alpha}(r_{p,t:t+h})$, is the solution of the following equation:*

$$e^{\top}\int_0^{\infty}\frac{\bar{D}_2(u)}{u}du\,\Pi-(\alpha-\frac{1}{2})\pi=0,$$

where Π represents the vector of the ergodic probabilities.

The h-period-ahead unconditional VaR is equal to the one-period-ahead unconditional VaR given by Corollary **3**. To compute the conditional or unconditional VaR of the h-period-ahead portfolio and aggregated portfolio one can follow the same steps of the algorithm described in subsection 3.3.1.

Proposition 8 (*Multi-Horizon Conditional Expected Shortfall*) *The h-period-ahead portfolio's conditional-Expected Shortfall with coverage probability α, denoted $ES^{\alpha}_{t+h}(r_{p,t+h})$, is given by:*

$$ES^{\alpha}_{t+h}(r_{p,t+h})=\frac{1}{\alpha\sqrt{2\pi}}R(u)P^{h-1}\Pi_{s_t}.$$

where

$$R(u)=Diag\left[\exp\left(-\frac{1}{2}\frac{(W^{\top}\mu_1+VaR(r_{p,t+h}))^2}{(W^{\top}\Omega_1 W)}\right),...,\exp(-\frac{1}{2}\frac{(W^{\top}\mu_N+VaR(r_{p,t+h}))^2}{(W^{\top}\Omega_N W)}\right)\right].$$

The h-period-ahead unconditional Expected Shortfall is equal to the one-period-ahead unconditional Expected Shortfall given by Corollary **4**.

3.4 Mean-Variance Efficient Frontier

In the literature there are two ways of considering the problem of portfolio optimization: static and dynamic. In the mean-variance framework, the difference between these two ways is related to how we calculate the first two moments of asset returns. In the static approach, the structure of the optimal portfolio is chosen once and for all at the beginning of the period. One critical drawback of this approach is that it assumes a constant mean and variance of returns. In the dynamic approach, the structure of the optimal portfolio is continuously adjusted using the available information set. One advantage of this approach is that it allows exploitation of the predictability of the first and second moments of asset returns to hedge changes in the investment opportunity set.

In this section and next one we study the multi-horizon portfolio optimization problem in the mean-variance context and under Markov switching model. We characterize the dynamic and static optimal portfolios and their term structure. This is to examine the relevance of risk horizon effects on the mean-variance efficient frontier and to compare the performance of the dynamic and static optimal portfolios.

3.4.1 Mean-Variance efficient frontier of dynamic portfolio

We consider risk-averse investors with preferences defined over the conditional (unconditional) expectation and variance-covariance matrix of portfolio returns. We provide a dynamic (static) frontier of all feasible portfolios characterized by a dynamic vector of weight W_t (a static vector of weight W). This frontier, which can be constructed from the n risky assets that we consider, is defined as the locus of feasible portfolios that have the smallest variance for a prescribed expected return.

The efficient frontier of dynamic portfolio can be described as the set of dynamic portfolios that satisfy the following constrained minimization problem

$$\begin{cases} \min\limits_{W_t \in \mathbb{W}} \frac{1}{2} \left\{ Var_t[r_{p,t+h}] = W_t^{\top} Var_t[r_{t+h}] W_t \right\} \\ \\ \text{st.} \\ \\ E_t[r_{p,t+h}] = W_t^{\top} E_t[r_{t+h}] = \bar{\mu}. \\ \\ W_t^{\top} e = 1. \end{cases} \tag{3.24}$$

where $\mathbb{W}$ is the set of all possible portfolios, $\bar{\mu}$ is the target expected return, and the mean $E_t[r_{t+h}]$ and variance $Var_t[r_{t+h}]$ are given in the following proposition.

Proposition 9 (*Multivariate Conditional Moments of Returns*) *The first and second conditional moments of the h-period-ahead multivariate return are given by:*

$$E_t[r_{t+h}] = E[r_{t+h} \mid I_t] = \bar{\mu}_t = \mu P^{h-1}\Pi_{s_t},\ h \geq 1,$$

$$Var_t[r_{t+h}] = Var[r_{t+h} \mid I_t] = \left(\Pi_{s_t}^{\top} \otimes I_n\right)\left(\left((P^{h-1})^{\top}\right) \otimes I_n\right)\bar{\Omega}_t,\ h \geq 2,$$

where

$$\bar{\Omega}_t = \begin{bmatrix} (\mu_1 - \bar{\mu}_t)\ (\mu_1 - \bar{\mu}_t)^{\top} + \Omega_1 \\ \cdots \\ (\mu_N - \bar{\mu}_t)(\mu_N - \bar{\mu}_t)^{\top} + \Omega_N \end{bmatrix},$$

I_n is an $n \times n$ identity matrix.

The Lagrangian of the minimizing problem (3.24) is given by:

$$L_t = \frac{1}{2}\left\{W_t^{\top} Var_t[r_{t+h}]W_t\right\} + \gamma_1\left\{\bar{\mu} - W_t^{\top}E_t[r_{t+h}]\right\} + \gamma_2\left\{1 - W_t^{\top}e\right\}. \qquad (3.25)$$

where γ_1 and γ_2 are the Lagrange multipliers. Under the first- and second-order conditions on the Lagrangian function (3.25), the solution of the above optimization problem is given by the following equation:

$$W_t^{opt} = \Lambda_1 + \Lambda_2\bar{\mu}. \qquad (3.26)$$

The $n \times 1$ vectors Λ_1 and Λ_2 are defined as follows:

$$\Lambda_1 = \tfrac{1}{A_4}[A_1 Var_t[r_{t+h}]^{-1}e - A_3 Var_t[r_{t+h}]^{-1}E_t[r_{t+h}]],$$
$$\Lambda_2 = \tfrac{1}{A_4}[A_2 Var_t[r_{t+h}]^{-1}E_t[r_{t+h}] - A_3 Var_t[r_{t+h}]^{-1}e]. \qquad (3.27)$$

where

$$
\begin{aligned}
A_1 &= E_t[r_{t+h}]^\top Var_t[r_{t+h}]^{-1} E_t[r_{t+h}], \\
A_2 &= e^\top Var_t[r_{t+h}]^{-1} e, \\
A_3 &= e^\top Var_t[r_{t+h}]^{-1} E_t[r_{t+h}], \\
A_4 &= A_1 A_2 - A_3^2.
\end{aligned}
\tag{3.28}
$$

The trading strategy implicit in equation (3.26) identifies the dynamically rebalanced portfolio with the lowest conditional variance for any choice of conditional expected return. From equations (3.26)-(3.28) it seems that forecasting future optimal weights requires on forecasting the expectation and variance of the portfolio's return. In the Markov switching regimes context, the first two moments can be predicted using Proposition 9. Many recent studies examine the economic implications of return predictability on investors' asset allocation decisions and found that investors react differently when returns are predictable. In the mean-variance framework, Jacobsen (1999) and Marquering and Verbeek (2001) found that the economic gains of exploiting return predictability are significant, whereas Handa and Tiwari (2004) found that the economic significance of return predictability is questionable.[1] In this chapter, we use a Markov switching model to examine the economic gains of return predictability. In the empirical application, we consider an ex ante analysis to compare the performance of the dynamic and static optimal portfolios. The measure of performance that we consider is given by the Sharpe ratio

$$
SR_t(W_t^{opt}) = \frac{W_t^{opt\top} E_t[r_{t+h}]}{\sqrt{W_t^{opt\top} Var_t[r_{t+h}] W_t^{opt}}}. \tag{3.29}
$$

If an investor believes that the conditional expected return and variance-covariance matrix of returns are constant, then the optimal weights will be constant over time-we refer

[1] For more discussion we refer the reader to Han (2005).

to them as static weights. The latter can be obtained by replacing the conditional first two moments given in Proposition 9 by those given in the the following proposition.

Proposition 10 (*Multivariate Unconditional Moments of Returns*) *The first and second unconditional moments of the h-period-ahead multivariate return are given by:*

$$E[r_{t+h}] = \bar{\mu} = \mu\Pi,\ h \geq 1,$$

$$Var[r_{t+h}] = (\Pi^{\top} \otimes I_n)\ (((P^{h-1})^{\top}) \otimes I_n)\bar{\Omega},\ h \geq 2,$$

where

$$\bar{\Omega} = \begin{bmatrix} (\mu_1 - \bar{\mu})\ (\mu_1 - \bar{\mu})^{\top} + \Omega_1 \\ \dots \\ (\mu_N - \bar{\mu})(\mu_N - \bar{\mu})^{\top} + \Omega_N \end{bmatrix}.$$

The static optimal portfolio allocation yields a constant Sharpe ratio, denoted $SR(W^{opt})$. In the empirical study, we compare the performance of the conditional and unconditional optimal portfolios by examining the proportion of times where

$$SR_t(W_t^{opt}) > SR(W^{opt}).$$

Finally, the relationship between $\bar{\mu}$ and the standard deviation of the optimal portfolio returns, denoted $\sigma_t^{opt}(r_{p,t+h})$, can be found from equation (3.26). It is characterized by the following equation:

$$\sigma_t^{opt}(r_{p,t+h}) = \sqrt{W_t^{opt\top} Var_t[r_{t+h}] W_t^{opt}} \tag{3.30}$$

which defines the mean-variance boundary, denoted $\boldsymbol{B(\mu, \sigma)}$. Equation (3.30) shows that there is a one-to-one relation between $B(\bar{\mu}, \sigma_t^{opt}(r_{p,t+h}))$ and the subset of optimal portfolios in $\mathbb{W}$. We have,

$$(\bar{\mu}, \sigma_t^{opt}(r_{p,t+h})) \in \boldsymbol{B}(\boldsymbol{\mu}, \boldsymbol{\sigma}) \Leftrightarrow \sigma_t^{opt}(r_{p,t+h}) = \sqrt{\frac{1}{A_2} + A_2\frac{(\bar{\mu} - A_3/A_2)^2}{A_4}}. \tag{3.31}$$

where the right-hand side of equation (3.31) defines a hyperbola in $\mathbb{R} \times \mathbb{R}^+$.

3.4.2 Term structure of the Mean-Variance efficient frontier

To study the term structure of the mean-variance efficient frontier, we consider the following optimization problem in which the efficient frontier at time t of the h-period-ahead aggregated portfolio can be described as the set of dynamic portfolios that satisfy the following constrained minimization problem

$$\begin{cases} \min_{\bar{W}_t \in \bar{\mathbb{W}}} \frac{1}{2}\{Var_t[r_{p,t:t+h}] = \bar{W}_t^\top Var_t[r_{t:t+h}]\bar{W}_t\} \\ \\ \text{st.} \\ \\ E_t[r_{p,t:t+h}] = \bar{W}_t^\top E_t[r_{t:t+h}] = \bar{\mu}. \\ \\ \bar{W}_t^\top e = 1. \end{cases} \tag{3.32}$$

where $\bar{\mathbb{W}}$ is the set of all possible portfolios, $\bar{\mu}$ is the target expected return, and the mean $E_t[r_{t:t+h}]$ and variance $Var_t[r_{t:t+h}]$ are given by the following proposition.

Proposition 11 (***Multivariate Conditional Moments of the Aggregated Returns***) *The first and second conditional moments of the h-period-ahead multivariate aggregated return are given by:*

$$E_t[r_{t:t+h}] = E[r_{t:t+h} \mid I_t] = \bar{\mu}_t = \mu[I + \sum_{l=1}^{h-1} P^l]\Pi_{s_t}, h \geq 2,$$

$$Var_t[r_{t:t+h}] = Var[r_{t:t+h} \mid I_t] = (\Pi_{s_t}^\top \otimes I_n)\bar{\Omega}_t + 2\mu(\sum_{l=1}^{h-1} P^l)Diag(\Pi_{s_t})[\mu - e^\top \otimes \bar{\mu}_t]^\top$$

$$+2\mu[\sum_{l=1}^{h-2} \sum_{k=1}^{h-l-1} P^k Diag(P^l \Pi_{s_t})]\mu^\top$$

$$+(\Pi_{s_t}^\top \otimes I_n)\ ((\sum_{l=1}^{h-1} P^l)^\top \otimes I_n)\Omega,\ h \geq 3.$$

where

$$\Omega = \begin{bmatrix} \mu_1\mu_1^\top + \Omega_1 \\ \cdots \\ \mu_N\mu_N^\top + \Omega_N \end{bmatrix},\ \bar{\Omega}_t = \begin{bmatrix} (\mu_1 - \bar{\mu}_t)\ (\mu_1 - \bar{\mu}_t)^\top + \Omega_1 \\ \cdots \\ (\mu_N - \bar{\mu}_t)(\mu_N - \bar{\mu}_t)^\top + \Omega_N \end{bmatrix}.$$

$$Diag(\Pi_{s_t}) = Diag\,(\mathsf{P}[s_t = 1 \mid I_t],\ldots,\mathsf{P}[s_t = N \mid I_t]).$$

The Lagrangian of the minimization problem (3.32) is given by:

$$L_t = \frac{1}{2}\{W_t^\top Var_t[r_{t:t+h}]W_t\} + \bar{\gamma}_1\{\bar{\mu} - W_t^\top E_t[r_{t:t+h}]\} + \bar{\gamma}_2\{1 - W_t^\top e\}, \tag{3.33}$$

where $\bar{\gamma}_1$ and $\bar{\gamma}_2$ are the Lagrange multipliers. Thus, under the first- and second-order conditions of the Lagrangian function (3.33), the solution to problem (3.32) is given by:

$$W_t^{opt} = \bar{\Lambda}_1 + \bar{\Lambda}_2\bar{\mu}. \tag{3.34}$$

The $n \times 1$ vectors $\bar{\Lambda}_1$ and $\bar{\Lambda}_2$ are defined as follows:

$$\bar{\Lambda}_1 = \frac{1}{\bar{A}_4}[\bar{A}_1 Var_t[r_{t:t+h}]^{-1}e - \bar{A}_3 Var_t[r_{t:t+h}]^{-1}E_t[r_{t:t+h}]],$$

$$\bar{\Lambda}_2 = \frac{1}{\bar{A}_4}[\bar{A}_2 Var_t[r_{t:t+h}]^{-1}E_t[r_{t:t+h}] - \bar{A}_3 Var_t[r_{t:t+h}]^{-1}e]$$

and

$$\bar{A}_1 = E_t[r_{t:t+h}]^\top Var_t[r_{t:t+h}]^{-1}E_t[r_{t:t+h}],$$

$$\bar{A}_2 = e^\top Var_t[r_{t:t+h}]^{-1}e,$$

$$\bar{A}_3 = e^\top Var_t[r_{t:t+h}]^{-1}E_t[r_{t:t+h}],$$

$$\bar{A}_4 = \bar{A}_1\bar{A}_2 - \bar{A}_3^2.$$

The unconditional weights of the aggregated portfolio simply follows from taking limits in (3.34) as $h \to \infty$. That is, we use the unconditional expectation and variance-covariance matrix of portfolio's returns implied by the Markov switching model (3.8).

Proposition 12 (***Multivariate Unconditional Moments of the Aggregated Returns***) *The first and second unconditional moments of the h-period-ahead multivariate aggregated return are given by:*

$$E[r_{t:t+h}] = \bar{\mu} = h\mu\Pi, h \geq 2,$$

$$Var[r_{t:t+h}] = (\Pi^\top \otimes I_n)\bar{\Omega} + 2\mu(\textstyle\sum_{l=1}^{h-1} P^l)Diag(\Pi)[\mu - e^\top \otimes \bar{\mu}]^\top$$

$$+2\mu[\textstyle\sum_{l=1}^{h-2}\sum_{k=1}^{h-l-1} P^k Diag(\Pi)]\mu^\top$$

$$+(\Pi^\top \otimes I_n)\ ((\textstyle\sum_{l=1}^{h-1} P^l)^\top \otimes I_n)\Omega,\ h \geq 3$$

where

$$\Omega = \begin{bmatrix} \mu_1\mu_1^\top + \Omega_1 \\ \ldots \\ \mu_N\mu_N^\top + \Omega_N \end{bmatrix}, \; \bar{\Omega} = \begin{bmatrix} (\mu_1 - \bar{\mu})\ (\mu_1 - \bar{\mu})\ ^\top + \Omega_1 \\ \ldots \\ (\mu_N - \bar{\mu})(\mu_N - \bar{\mu})^\top + \Omega_N \end{bmatrix}, \; Diag(\Pi) = Diag\,(\pi_1,\ldots,\pi_N).$$

An investor who uses the dynamic optimization approach will perceive the risk-return trade-off differently than an investor who uses the static approach. With the dynamic optimization approach we will have a different return expectation and risk (variance) each period. Long-term risks of asset returns may differ from their short-term risks. In the static approach, the variance of each asset return is proportional to the horizon over which it is held. It is independent of the time horizon, and thus there is a single number that summarizes risks for all holding periods [see Campbell and Viceira (2005)]. In the dynamic optimization approach model, by contrast, the variance may either increase or decline as the holding period increases [see our empirical results].

The relationship between $\bar{\mu}$ and the standard deviation of the optimal portfolio return, denoted $\bar{\sigma}_t^{opt}(r_{p,t:t+h})$, can be found from equation (3.34). This is characterized by the following equation:

$$\bar{\sigma}_t^{opt}(r_{p,t:t+h}) = \sqrt{\bar{W}_t^{opt\top} Var_t[r_{t:t+h}]\bar{W}_t^{opt}},$$

which defines the mean-variance boundary, denoted $\bar{\boldsymbol{B}}(\boldsymbol{\mu},\boldsymbol{\sigma})$. Equation (3.34) shows that there is a one-to-one relation between $\bar{B}(\bar{\mu},\ \bar{\sigma}_t^{opt}(r_{p,t+h}))$ and the subset of optimal portfolios in $\bar{\mathbb{W}}$.We have,

$$(\bar{\mu}, \bar{\sigma}_t^{opt}(r_{p,t:t+h})) \in \bar{\boldsymbol{B}}(\boldsymbol{\mu},\boldsymbol{\sigma}) \Leftrightarrow \bar{\sigma}_t^{opt}(r_{p,t:t+h}) = \sqrt{\frac{1}{\bar{A}_2} + \bar{A}_2\frac{(\bar{\mu} - \bar{A}_3/\bar{A}_2)^2}{\bar{A}_4}}, \qquad (3.35)$$

where the right-hand side of equation (3.35) defines a hyperbola in $\mathbb{R} \times \mathbb{R}^+$.

3.5 Empirical Application

In this section, we use real data (Standard and Poor's and Toronto Stock Exchange Composite indices) to examine the impact of the asset return predictability on the variance and VaR of linear portfolio across investment horizons. We analyze the relevance of risk horizons effects on the multi-horizon mean-variance efficient frontier and we compare the performance of the dynamic and static optimal portfolios.

3.5.1 Data and parameter estimates

Our data consists of daily observations on prices from S&P 500 and TSE 300 indices contracts from January 1988 through May 1999, totalling 2959 trading days. The asset returns are calculated by applying the standard continuous compounding formula, $r_{it} = 100\times\ (ln(P_{it})\ - ln(P_{it-1}))$ for $i = 1, 2$, where P_{it} is the price of the asset i. Summary statistics for S&P 500 and TSE 300 daily returns are given in Tables 4 and 5. These daily returns are displayed in Figures 1 and 2. Looking at these tables and figures, we note some main stylized facts. The unconditional distributions of S&P 500 and TSE 300 daily returns show the expected excess kurtosis and negative skewness. The sample kurtosis is much greater than the normal distribution value of three for all series. The time series plots of these daily returns show the familiar volatility clustering effect, along with a few occasional very large absolute returns.

To implement the results of the previous sections, we consider two-state bivariate Markov switching model. The results of the estimation of this model are given in Table 6. We see that there are significant time-variations in the first and second moments of the joint distribution of the S&P 500 and TSE 300 stocks across the two regimes. Mean returns to the S&P 500 stock vary from 0.0890 per day in the first state to -0.0327 per day in the second state. Mean returns to the TSE 300 vary from 0.0738 per day in the first state to -0.1118 per day in the second state. All estimates of mean stock returns are statistically significant except for the S&P 500 in the second state. For the volatility and correlation parameters, we found that S&P 500 stock return volatility varies between 0.4098 and almost 2.0895 per day, with state two displaying the highest value. TSE 300 stock returns show less variation and their volatility varies between 0.2039 and 1.4354 per day, with more volatility in state two. Correlations between S&P 500 and TSE 300 returns vary between 0.5584 in state one and 0.7306 in state two. Finally, the transition probability estimates and the smoothed and filtered state probability plots [see Figures 3 and 4] reveal that states one and two capture 80% and 20% of the sample, respectively, implying that regime one is more persistent than regime two.

3.5.2 Results

We examine the implications of asset return predictability for risk (variance and VaR) across investment horizons. We analyze the relevance of risk horizon effects on the mean-variance efficient frontier and we compare the performance of the dynamic and static optimal portfolios. We present our empirical results mainly through graphs.

By considering linear portfolio on S&P 500 and TSE 300 indices, we find that the multi-horizon conditional variance of the optimal portfolio is time-varying and shows the familiar volatility clustering effect [see Figures 5-7]. This proves the ability of the Markov switching model to account for volatility clustering observed in the stock prices. At a given point in time[5] t and when we lengthen the horizon h, Figure 8 shows the convergence of the conditional variance to the unconditional one. The conditional variance per period of the multi-horizon optimal portfolio's returns, when plotted as a function of the horizon h may be increasing or decreasing at intermediate horizons, and it eventually converges to a constant-the unconditional variance-at long enough horizons.[6]

Figures 9-11 show that the conditional 5% VaR of the optimal portfolio is time-varying and persistent [see Engle and Manganelli (2002)]. At a given point in time t, Figure 12 shows that the conditional VaR converges to the unconditional one. The latter is given

[5]For illustration we take $t = 680$, $t = 1000$, and $t = 2958$.

[6]This result is similar to the one found in Campbell and Viceira (2005).

by a flat line and means that the level of risk is the same at short and long horizons. However, the conditional VaR may increases or decreases with horizons depending on the point in time where we are. For example, at $t = 680$ and $t = 1000$, the conditional VaR decreases with the horizon and it is bigger than the unconditional VaR [see Figure 12]. Consequently, considering only unconditional VaR may under- or overestimate risk across investment horizons. Same results hold for the 10% VaR [see Figure 13].

Figures 15-17 show that the conditional mean-variance efficient frontier is time-varying and converges to the unconditional efficient frontier given in Figure 14. When the multi-horizon expected returns and risk (variance) are flat, the efficient frontier is the same at all horizons, and short-term mean-variance analysis provides answers that are valid for all mean-variance investors, regardless of their investment horizon. However, when the multi-horizon expected returns and risk are time-varying, efficient frontiers at different horizons may not coincide. In that case short-term mean-variance analysis can be misleading for investors with longer investment horizons. The above results are similar to these found by Campbell and Viceira (2005) and may confirmed by Figures 18-20 who show that the conditional Sharpe ratio of the optimal portfolio is time-varying and converges to a constant- the unconditional Sharpe ratio. Figures 18-20, also show the presence of clustering phenomena in the conditional Sharp ratio. At a given point in time t the conditional mean-variance frontier may be more efficient than the unconditional one [see Figure 15]. To check the latter result, we compare the performance of the conditional and unconditional optimal portfolios. We look at the proportion of times where the conditional one period-ahead Sharpe ratio is bigger than the unconditional one:

$$SR_t(W_t^{opt}) > SR(W^{opt}).$$

and the empirical results show that in 73.56% of the sample the conditional optimal portfolio performs better then the unconditional one.

For the aggregated optimal portfolio, we first note the time-varying and volatility clustering effect in the conditional variance [see Figures 23-24]. The conditional and unconditional variances increase with the horizon h [see Figure 25]. Specially, the unconditional variance is a linear increasing function of h. and this means that the unconditional variances are independent of the time horizon, and thus there is a single number that summarizes risks for all holding periods. Second, it seems from Figures 27-29 that the conditional 5 % VaR is time-varying and it is a non-linear increasing function of the horizon h. Figure 29, shows that the conditional 5 % VaR may be bigger or smaller than the unconditional one depending on the point in time where we are. For $t = 680$ and $t = 1000$, we see that the unconditional 5 % VaR underestimates risk, since it is smaller than the conditional 5 % VaR. Again, considering only unconditional VaR may under- or overestimate risk in the aggregated optimal portfolio across investment horizons. Same results holds for the 10% VaR [see Figure 30]. Finally, Figures 31-34, show that the conditional and unconditional mean-variance frontiers of the aggregated optimal portfolio become large and more efficient when we increase the horizon h. These results are confirmed by Figure 38 who show that the conditional and unconditional Sharpe ratios increase with the horizon h.

3.6 Conclusion

In this chapter, we consider a Markov switching model to capture important features such as heavy tails, persistence, and nonlinear dynamics in the distribution of asset returns. We compute the conditional probability distribution function of the multi-horizon portfolio's returns, which we use to approximate the conditional Value-at-Risk (VaR). We derive a closed-form solution for the multi-horizon conditional Expected Shortfall and we characterize the multi-horizon mean-variance efficient frontier of the optimal portfolio. Using daily observations on S&P 500 and TSE 300 indices, we first found that the conditional risk (variance and VaR) per period of the multi-horizon optimal portfolio's returns, when plotted as a function of the horizon, may be increasing or decreasing at intermediate horizons, and converges to a constant- the unconditional risk-at long enough horizons. Second, the efficient frontiers of the multi-horizon optimal portfolios are time-varying. Finally, at short-term and in 73.56% of the sample the conditional optimal portfolio performs better then the unconditional one.

3.7 Appendix: Proofs

Appendix 1: Proof of Propositions

Proof of Proposition 1. We have

$$
\begin{aligned}
E[\exp(u^\top \zeta_{t+1}) \mid J_t] &= \sum_{i=1}^{N} \exp(u_i) \mathsf{P}(s_{t+1} = i \mid J_t) \\
&= (\exp(u_1), ..., \exp(u_N)) \;\; (\mathsf{P}(s_{t+1} = 1 \mid J_t) ... \mathsf{P}(s_{t+1} = N \mid J_t))^\top \\
&= e^\top Diag(\exp(u_1), ..., \exp(u_N)) \;\; P\zeta_t \\
&= e^\top A(u)\zeta_t. \qquad (3.36)
\end{aligned}
$$

Therefore, for $h \geq 2$

$$
\begin{aligned}
E[\exp(u^\top \zeta_{t+h}) \mid J_t] &= E[e^\top A(u)\zeta_{t+h-1} \mid J_t] \\
&= e^\top A(u) E[\zeta_{t+h-1} \mid J_t] \\
&= e^\top A(u) P^{h-1} \zeta_t.
\end{aligned}
$$

where the last equality follows from (3.1). Similarly,

$$
\begin{aligned}
E[\exp(u^\top \zeta_{t+1})\zeta_{t+1} \mid J_t] &= \sum_{i=1}^{N} \exp(u_i) \mathsf{P}(s_{t+1} = i \mid J_t) e_i \\
&= Diag(\exp(u_1), ..., \exp(u_N))(\mathsf{P}(s_{t+1} = 1 \mid J_t), ..., \mathsf{P}(s_{t+1} = N \mid J_t))^\top \\
&= Diag(\exp(u_1), ..., \exp(u_N)) \;\; P\zeta_t \\
&= A(u)\zeta_t. \qquad (3.37)
\end{aligned}
$$

Observe that one gets (3.36) from (3.37) by multiplying it by $e^\top$:

$$
e^\top E[\exp(u^\top \zeta_{t+1})\zeta_{t+1} \mid J_t] = E[\exp(u^\top \zeta_{t+1}) e^\top \zeta_{t+1} \mid J_t] = E[\exp(u^\top \zeta_{t+1}) \mid J_t].
$$

given that $e^\top \zeta_{t+1} = 1.$■

Proof of Proposition 2.

$$
\begin{aligned}
E[\exp\left(\sum_{i=1}^{h} u_i^\top \zeta_{t+i}\right) \mid J_t] &= E[\exp\left(\sum_{i=1}^{h-1} u_i^\top \zeta_{t+i}\right) E[\exp(u_h^\top \zeta_{t+h}) \mid J_{t+h-1}] \mid J_t] \\
&= E[\exp\left(\sum_{i=1}^{h-1} u_i^\top \zeta_{t+i}\right) e^\top A(u_h)\zeta_{t+h-1} \mid J_t] \\
&= e^\top A(u_h) E[\exp\left(\sum_{i=1}^{h-1} u_i^\top \zeta_{t+i}\right) \zeta_{t+h-1} \mid J_t] \\
&= e^\top A(u_h) E[\exp\left(\sum_{i=1}^{h-2} u_i^\top \zeta_{t+i}\right) E[\exp(u_{h-1}^\top \zeta_{t+h-1})\zeta_{t+h-1} \mid J_{t+h-2}] \mid J_t] \\
&= e^\top A(u_h) A(u_{h-1}) E[\exp\left(\sum_{i=1}^{h-2} u_i^\top \zeta_{t+i}\right) \zeta_{t+h-2} \mid J_t].
\end{aligned}
$$

By iterating the last two equalities, one gets (3.6). By taking the unconditional expectation of (3.6) and by using (3.3), one gets (3.7).■

Proof of Proposition 3. Given the information set J_t, the distribution of r_{t+1} is $\mathcal{N}(\mu\zeta_t, \Omega(\zeta_t))$. Thus, $\forall\ U = (u_1, ..., u_n) \in \mathbb{R}^n$, we have

$$
E[\exp(iU^\top r_{t+1}) \mid J_t] = \exp\left(iU^\top \mu\zeta_t - \frac{U^\top \Omega(\zeta_t) U}{2}\right).
$$

Observe that,

$$
U^\top \Omega(\zeta_t) U \quad = \quad \sum_{l_1=1}^{n} \sum_{l_2=1}^{n} u_{l_1}\ u_{l_2} \omega_{l_1 l_2}^\top \zeta_t = \left(\sum_{1 \le l_1, l_2 \le n} u_{l_1} u_{l_2} \omega_{l_1 l_2} \right)^\top \zeta_t.
$$

If we take $U = u\,(\alpha_1, \alpha_2, ..., \alpha_n)^\top$, then the characteristic function of $r_{p,t+1}$ is

$$
E[\exp(iur_{p,t+1}) \mid J_t] = \exp\left(\left(iu\mu^\top W - \frac{u^2}{2} \sum_{1 \le l_1, l_2 \le n} \alpha_{l_1} \alpha_{l_2} \omega_{l_1 l_2}\right)^\top \zeta_t\right).
$$

i.e., (3.10). ■

Proof of Proposition 5.

$$\begin{aligned} ES_t^{\alpha}(r_{p,t+1}) &= E_t[r_{p,t+1} \mid r_{p,t+1} \leq -VaR_t^{\alpha}(r_{p,t+1})] \\ &= \int_{-\infty}^{-VaR_t^{\alpha}(r_{p,t+1})} r_p \, f_t(r_p \mid r_p \leq -VaR_t^{\alpha}(r_{p,t+1}))dr_p \\ &= \int_{-\infty}^{-VaR_t^{\alpha}(r_{p,t+1})} r_p \, \frac{\sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t)\frac{1}{\sqrt{2\pi(W^{\top}\Omega_j W)}}\exp(-\frac{1}{2}\frac{(r_p - W^{\top}\mu_j)^2}{(W^{\top}\Omega_j W)})}{\mathsf{P}_t(r_p \leq -VaR_t^{\alpha}(r_{p,t+1}))} dr_p. \end{aligned}$$

Since $\mathsf{P}_t(r_p \leq -VaR_t^{\alpha}(r_{p,t+1})) = \alpha$, we have

$$\begin{aligned} ES_t^{\alpha}(r_{p,t+1}) &= \frac{1}{\alpha\sqrt{2\pi}}\sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t)\int_{-\infty}^{-VaR_t^{\alpha}(r_{p,t+1})} \frac{r_p}{\sqrt{(W^{\top}\Omega_j W)}}\exp\left(-\frac{1}{2}\frac{(r_p - W^{\top}\mu_j)^2}{(W^{\top}\Omega_j W)}\right) dr_p \\ &= \frac{1}{\alpha\sqrt{2\pi}}\sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t)\exp\left(-\frac{1}{2}\frac{(W^{\top}\mu_j + VaR_t^{\alpha}(r_{p,t+1}))^2}{(W^{\top}\Omega_j W)}\right). \end{aligned}$$

$ES_t^{\alpha}(r_{p,t+1})$ can be written as follows:

$$ES_{t+1}^{\alpha}(r_{p,t+1}) = \frac{1}{\alpha\sqrt{2\pi}} e^{\top}\mathrm{R}(u)\Pi_{s_t},$$

where

$$\mathrm{R}(u) = Diag\left[\exp\left(-\frac{1}{2}\frac{(W^{\top}\mu_1 + VaR_t(r_{p,t+1}))^2}{(W^{\top}\Omega_1 W)}\right),\ldots\ldots,\exp\left(-\frac{1}{2}\frac{(W^{\top}\mu_N + VaR_t(r_{p,t+1}))^2}{(W^{\top}\Omega_N W)}\right)\right].$$

■

Proof of Proposition 6. Given the information set $J_t^* = J_t \cup \{s_{t+h-1}\}$, we have

$$r_{t+h} \mid J_t^* \sim \mathcal{N}(\mu\zeta_{t+h-1}, \Omega(\zeta_{t+h-1})).$$

Consequently, $\forall\, U = (u_1, ..., u_n) \in \mathbb{R}^n$,

$$\begin{aligned} E[\exp(iU^\top r_{t+h}) \mid J_t] &= E[E[\exp(iU^\top r_{t+h}) \mid J_t^*] \mid J_t] \\ &= E\left[\exp\left(iU^\top \mu\zeta_{t+h-1} - \frac{1}{2}U^\top \Omega(\zeta_{t+h-1})U\right) \mid J_t\right] \\ &= E\left[\exp\left(\left(i\mu^\top U - \frac{1}{2}\sum_{1\le l_1,l_2\le n} u_{l_1}u_{l_2}\omega_{l_1l_2}\right)^\top \zeta_{t+h-1}\right) \mid J_t\right]. \end{aligned}$$

Using Proposition **1**,we get

$$E[\exp(iU^\top r_{t+h}) \mid J_t] = e^\top A\left(i\mu^\top U - \frac{1}{2}\sum_{1\le l_1,l_2\le n} u_{l_1}u_{l_2}\omega_{l_1l_2}\right)P^{h-2}\zeta_t.$$

where

$$\mathrm{A}\left(i\mu^\top U - \frac{1}{2}\sum_{1\le l_1,l_2\le n} u_{l_1}u_{l_2}\omega_{l_1l_2}\right) = Diag\left(\exp\left(iU^\top\mu_1 - \frac{1}{2}U^\top\Omega_1 U\right), ..., \exp\left(iU^\top\mu_N - \frac{1}{2}U^\top\Omega_N U\right)\right).$$

If we let $U = u(\alpha_1, \alpha_2, ..., \alpha_n)^\top$, then the conditional characteristic function of portfolio's returns, $r_{p,t+h}$, is given by:

$$E[\exp(iur_{p,t+h}) \mid J_t] = e^\top A(iu\mu^\top W - \frac{u^2}{2}\sum_{1\le l_1,l_2\le n} \alpha_{l_1}\alpha_{l_2}\omega_{l_1l_2})P^{h-2}\zeta_t,$$

i.e., (3.20). Similarly, from (3.8)

$$r_{t:t+h} = \sum_{k=1}^{h}\left(\mu\zeta_{t+k-1} + \Sigma(\zeta_{t+k-1})\varepsilon_{t+k}\right), \quad \varepsilon_{t+k}\ i.i.d. \sim \mathcal{N}(0, I_n). \qquad (3.38)$$

Given the information set $J_t^{**} = J_t \cup \{s_{t+1}, \ldots, s_{t+h-1}\}$, we have

$$r_{t:t+h} \mid J_t^{**} \sim \mathcal{N}\left(\sum_{k=1}^{h} \mu\zeta_{t+k-1}, \sum_{k=1}^{h} \Omega(\zeta_{t+k-1})\right).$$

Consequently, $\forall\, U = (u_1, \ldots, u_n) \in \mathbb{R}^n$,

$$\begin{aligned}
E[\exp(iU^\top r_{t:t+h}) \mid J_t] &= E[E[\exp(iU^\top r_{t:t+h}) \mid J_t^{**}] \mid J_t] \\
&= E[\exp\left(iU^\top \sum_{k=1}^{h} \mu\zeta_{t+k-1} - \frac{1}{2}\sum_{k=1}^{h} U^\top \Omega(\zeta_{t+k-1})U\right) \mid J_t] \\
&= E[\exp\left(\sum_{k=1}^{h} iU^\top \mu\zeta_{t+k-1} - \frac{1}{2}U^\top \Omega(\zeta_{t+k-1})U\right) \mid J_t] \\
&= E[\exp\left(\sum_{k=1}^{h}\left(i\mu^\top U - \frac{1}{2}\sum_{1\le l_1,l_2\le n} u_{l_1}u_{l_2}\omega_{l_1l_2}\right)^\top \zeta_{t+k-1}\right) \mid J_t] \\
&= E[\exp\left(\sum_{k=1}^{h-1}\left(i\mu^\top U - \frac{1}{2}\sum_{1\le l_1,l_2\le n} u_{l_1}u_{l_2}\omega_{l_1l_2}\right)^\top \zeta_{t+k}\right) \mid J_t] \\
&\quad \times \exp\left(\left(i\mu^\top U - \frac{1}{2}\sum_{1\le l_1,l_2\le n} u_{l_1}u_{l_2}\omega_{l_1l_2}\right)^\top \zeta_t\right).
\end{aligned}$$

Using Proposition **2**, we get

$$\mathrm{E}[\exp\left(\sum_{k=1}^{h-1}\left(i\mu^\top U - \frac{1}{2}\sum_{1\le l_1,l_2\le n} u_{l_1}u_{l_2}\omega_{l_1l_2}\right)^\top \zeta_{t+k}\right) |\mathrm{J}_t] = e^\top\left(A\left(i\mu^\top U - \frac{1}{2}\sum_{1\le l_1,l_2\le n} u_{l_1}u_{l_2}\omega_{l_1l_2}\right)\right)^{h-1}\zeta_t.$$

and

$$\begin{aligned}
E[\exp(iU^\top r_{t:t+h}) \mid J_t] &= e^\top(A(i\mu^\top U - \frac{1}{2}\sum_{1\le l_1,l_2\le n} u_{l_1}u_{l_2}\omega_{l_1l_2}))^{h-1} \\
&\quad \times \exp((i\mu^\top U - \frac{1}{2}\sum_{1\le l_1,l_2\le n} u_{l_1}u_{l_2}\omega_{l_1l_2})^\top\zeta_t)\zeta_t.
\end{aligned}$$

If we let $U = u(\alpha_1, \alpha_2, ..., \alpha_n)^\top$, then the conditional characteristic function of the aggregated portfolio's return, $r_{p,t:t+h}$, is given by:

$$E[\exp(iur_{p,t:t+h}) \mid J_t] = e^\top \left(A\left(iu\mu^\top W - \frac{u^2}{2}\sum_{1\leq l_1,l_2\leq n}\alpha_{l_1}\alpha_{l_2}\omega_{l_1 l_2}\right)\right)^{h-1}$$

$$\times \exp\left(\left(iu\mu^\top W - \frac{u^2}{2}\sum_{1\leq l_1,l_2\leq n}\alpha_{l_1}\alpha_{l_2}\omega_{l_1 l_2}\right)^\top \zeta_t\right)\zeta_t,$$

i.e., (3.21). ■

Proof of Proposition 8. Same proof as in Proposition 5. ■

Proof of Proposition 9. Given the constant $\bar{E} \in \mathbb{R}$, we have

$$\phi_{t+h}(u) = E[\exp(iu(r_{p,t+h} - \bar{E})) \mid I_t] = e^\top \bar{A}(u)P^{h-2}\Pi_{s_t},$$

where, $\forall u \in \mathbb{R}$,

$$\begin{aligned}\bar{A}(u) &= A\left(iu(\mu^\top W - \bar{E}\ e) - \frac{u^2}{2}\sum_{1\leq l_1,l_2\leq n}\alpha_{l_1}\alpha_{l_2}\omega_{l_1 l_2}\right) \\ &= Diag\left(\exp\left(iu(W^\top\mu_1 - \bar{E}) - \frac{u^2}{2}(W^\top\Omega_1 W)\right), ..., \exp\left(iu(W^\top\mu_N - \bar{E}) - \frac{u^2}{2}(W^\top\Omega_N W)\right)\right)\end{aligned}$$

The first derivative of $\phi_{t+h}(u)$ with respect to u is given by:

$$\frac{d\phi_{t+h}(u)}{du} = e^\top\frac{d\bar{A}(u)}{du}P^{h-2}\Pi_{s_t} = e^\top\tilde{B}(u)P^{h-2}\Pi_{s_t},$$

where

$$\tilde{\mathrm{B}}(u) = Diag(\tilde{\mathrm{B}}(u)_1, ..., \tilde{\mathrm{B}}(u)_N)\mathrm{P}.$$

and, for $j = 1, ...N,$

$$\tilde{\mathrm{B}}(u)_j = \left[i(W^\top\mu_j - \bar{E}) - uW^\top\Omega_j W\right] \exp\left(iu(W^\top\mu_j - \bar{E}) - \frac{u^2}{2}W^\top\Omega_j W\right).$$

Consequently,

$$\frac{d\phi_{t+h}(0)}{du} = ie^\top \bar{B}(0)P^{h-2}\Pi_{s_t}.$$

where

$$\bar{B}(0) = Diag\left((W^\top\mu_1 - \bar{E}), ..., (W^\top\mu_N - \bar{E})\right) P. \tag{3.39}$$

For $\bar{E} = 0$, we get

$$E_t[r_{p,t+h}] \quad = \quad \frac{\phi^{(1)}_{t+h}(0)}{i} = e^\top B(0)P^{h-2}\Pi_{s_t} = W^\top\mu P^{h-1}\Pi_{s_t}.$$

Now, let us calculate the variance of $r_{p,t+h}$. Setting $\bar{E} = E_t[r_{p,t+h}]$.

$$\phi^{(2)}_{t+h}(u) = e^\top \frac{d\tilde{B}(u)}{du} P^{h-2}\Pi_{s_t} = e^\top \tilde{C}(u)P^{h-2}\Pi_{s_t}.$$

where,

$$\tilde{C}(u) = Diag(\tilde{C}_1(u), ..., \tilde{C}_N(u))P.$$

and, for $j = 1, ...N,$

$$\mathrm{C}_j(u) = \exp\left(iu(\mathrm{W}^\top\mu_j - \bar{E}) - \frac{u^2}{2}(\mathrm{W}^\top\Omega_j\mathrm{W})\right)\left[(i(\mathrm{W}^\top\mu_j - \bar{E}) - u\mathrm{W}^\top\Omega_j\mathrm{W})^2 - (\mathrm{W}^\top\Omega_j\mathrm{W})\right].$$

Consequently,

$$\phi^{(2)}_{t+h}(0) = i^2 e^\top \tilde{C}(0)P^{h-2}\Pi_{s_t}.$$

where, $\forall u \in \mathbb{R}$.

$$\tilde{C}(0) = Diag\left((W^\top\mu_1 - \bar{E})^2 + W^\top\Omega_1 W, ..., (W^\top\mu_N - \bar{E})^2 + W^\top\Omega_N W\right) P.$$

For $\bar{E} = W^\top \bar{\mu}_t$, we get

$$Var_t(r_{p,t+h}) = \frac{\phi^{(2)}_{t+h}(0)}{i^2} = W^\top \left[\left(\Pi_{s_t}^\top \otimes I_n \right) \left((P^{h-1})^\top \otimes I_n \right) \bar{\Omega}_t \right] W,$$

where

$$\bar{\Omega}_t = \begin{bmatrix} (\mu_1 - \bar{\mu}_t)(\mu_1 - \bar{\mu}_t)^\top + \Omega_1 \\ \dots \\ (\mu_N - \bar{\mu}_t)(\mu_N - \bar{\mu}_t)^\top + \Omega_N \end{bmatrix}, \ \bar{\mu}_t = \mu P^{h-1} \Pi_{s_t}.$$

■

Proof of Proposition 10. Proposition 10 can be deduced from Proposition 9 using the law of iterated expectations. ■

Proof of Proposition 11. Given the constant $\bar{E} \in \mathbb{R}$, we have

$$\begin{aligned} \phi_{t:t+h}(u) &= E[\exp\left(iu(r_{p,t:t+h} - \bar{E})\right) \mid I_t] \\ &= \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \exp\left(iu(W^\top \mu_j - \bar{E}) - \frac{u^2}{2}(W^\top \Omega_j W)\right) \left(e^\top \bar{A}(u)^{h-1} e_j\right). \end{aligned}$$

where, $\forall u \in \mathbb{R}$,

$$\bar{A}(u) = Diag\left(\exp\left(iuW^\top \mu_1 - \frac{u^2}{2}(W^\top \Omega_1 W)\right), \dots, \exp\left(iuW^\top \mu_N - \frac{u^2}{2}(W^\top \Omega_N W)\right)\right) P,$$

e_j is an $N \times 1$ vector of zeros with a one as its j^{th} element. The first derivative of $\phi_{t:t+h}(u)$

with respect to u is given by:

$$\frac{d\phi_{t:t+h}(u)}{du} = \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \frac{d}{du} \left\{ \exp\left(iu(W^\top \mu_j - \bar{E}) - \frac{u^2}{2}(W^\top \Omega_j W) \right) \left(e^\top \bar{A}(u)^{h-1} e_j \right) \right\}$$

$$= \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \left\{ \left(i(W^\top \mu_j - \bar{E}) \right) \exp\left(iu(W^\top \mu_j - \bar{E}) - \frac{u^2}{2}(W^\top \Omega_j W) \right) \left(e^\top \bar{A}(u)^{h-1} e_j \right) \right\}$$

$$+ \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \left\{ \left(-u(W^\top \Omega_j W) \right) \exp\left(iu(W^\top \mu_j - \bar{E}) - \frac{u^2}{2}(W^\top \Omega_j W) \right) \left(e^\top \bar{A}(u)^{h-1} e_j \right) \right.$$

$$+ \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \left\{ \exp\left(iu(W^\top \mu_j - \bar{E}) - \frac{u^2}{2}(W^\top \Omega_j W) \right) \left(e^\top \frac{d\bar{A}(u)^{h-1}}{du} e_j \right) \right\}$$

where

$$\frac{d\bar{A}(u)^{h-1}}{du} = \sum_{l=0}^{h-2} \bar{A}(u)^{h-2-l} \frac{d\bar{A}(u)}{du} \bar{A}(u)^l = \sum_{l=0}^{h-2} \bar{A}(u)^{h-2-l} B(u) \bar{A}(u)^l,$$

$$B(u) = Diag(B(u)_1, \ldots, B(u)_N) P,$$

and, for $j = 1, \ldots, N$,

$$B(u)_j = \left(iW^\top \mu_j - uW^\top \Omega_j W \right) \exp\left(iuW^\top \mu_j - \frac{u^2}{2} W^\top \Omega_j W \right).$$

For $\bar{E} = 0$, we get

$$E_t[r_{p,t:t+h}] = \frac{\phi^{(1)}_{t:t+h}(0)}{i} = \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \left\{ (W^\top \mu_j) \left(e^\top \bar{A}(0)^{h-1} e_j \right) \right\}$$

$$+ \frac{1}{i} \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \left\{ e^\top \frac{d\bar{A}(u)^{h-1}}{du} e_j \,\Big|_{u=0} \right\}.$$

Observe that, $\forall j = 1, ..., N$ and $\forall h \geq 0$,

$$\frac{d\bar{A}(u)^{h-1}}{du} \mid_{u=0} = i \sum_{l=0}^{h-2} P^{h-2-l} \bar{B}(0) P^{l},$$

$$\bar{A}(0)^{h-1} = P^{h-1},$$

$$e^{\top} \bar{A}(0)^{h-1} e_j = 1, \quad e^{\top} P^{h} = e^{\top},$$

where,

$$\bar{B}(0) = Diag(W^{\top} \mu_1, ..., W^{\top} \mu_N) P. \tag{3.40}$$

Consequently,

$$\begin{aligned}
E_t[r_{p,t:t+h}] &= W^{\top} \mu \Pi_{s_t} + \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \left(e^{\top} \sum_{l=0}^{h-2} P^{h-2-l} \bar{B}(0) P^{l} e_j \right) \\
&= W^{\top} \mu \Pi_{s_t} + \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \left(\sum_{l=0}^{h-2} W^{\top} \mu P^{l+1} e_j \right) \\
&= W^{\top} \mu \Pi_{s_t} + W^{\top} \mu \left(\sum_{l=0}^{h-2} P^{l+1} \right) \Pi_{s_t} \\
&= W^{\top} \mu \left[I + \sum_{l=1}^{h-1} P^{l} \right] \Pi_{s_t}.
\end{aligned}$$

Now, let us calculate the variance of $r_{p,t:t+h}$. Setting $\bar{E} = E_t[r_{p,t:t+h}]$

$$\phi^{(2)}_{t:t+h}(u) = \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \tfrac{d}{du} \{ \left(i(W^\top \mu_j - \bar{E}) - u(W^\top \Omega_j W) \right)$$

$$\times \exp\left(iu(W^\top \mu_j - \bar{E}) - \tfrac{u^2}{2}(W^\top \Omega_j W) \right) \left(e^\top \bar{A}(u)^{h-1} e_j \right) \}$$

$$+ \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \tfrac{d}{du} \left\{ \exp\left(iu(W^\top \mu_j - \bar{E}) - \tfrac{u^2}{2}(W^\top \Omega_j W) \right) \left(e^\top \tfrac{d\bar{A}(u)^{h-1}}{du} e_j \right) \right\}$$

$$= \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \exp\left(iu(W^\top \mu_j - \bar{E}) - \tfrac{u^2}{2}(W^\top \Omega_j W) \right)$$

$$\times \left\{ \left[\left(i(W^\top \mu_j - \bar{E}) - u(W^\top \Omega_j W) \right)^2 - (W^\top \Omega_j W) \right] \left(e^\top \bar{A}(u)^{h-1} e_j \right) \right\}$$

$$+ 2 \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \exp\left(iu(W^\top \mu_j - \bar{E}) - \tfrac{u^2}{2}(W^\top \Omega_j W) \right)$$

$$\times \left\{ \left(i(W^\top \mu_j - \bar{E}) - u(W^\top \Omega_j W) \right) \left(e^\top \tfrac{d\bar{A}(u)^{h-1}}{du} e_j \right) \right\}$$

$$+ \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \exp\left(iu(W^\top \mu_j - \bar{E}) - \tfrac{u^2}{2}(W^\top \Omega_j W) \right) \left(e^\top \tfrac{d^2\bar{A}(u)^{h-1}}{(du)^2} e_j \right),$$

where

$$\begin{aligned}
\frac{d^2 \bar{A}(u)^{h-1}}{(du)^2} &= \frac{d}{du} \left[\sum_{l=0}^{h-2} \bar{A}(u)^{h-2-l} \frac{d\bar{A}(u)}{du} \bar{A}(u)^l \right] \\
&= \sum_{l=0}^{h-2} \frac{d\bar{A}(u)^{h-2-l}}{du} B(u) \bar{A}(u)^l \\
&\quad + \sum_{l=0}^{h-2} \bar{A}(u)^{h-2-l} C(u) \bar{A}(u)^l \\
&\quad + \sum_{l=0}^{h-2} \bar{A}(u)^{h-2-l} B(u) \frac{d\bar{A}(u)^l}{du}.
\end{aligned}$$

$$C(u) = Diag(C_1(u), \ldots, C_N(u))\ P.$$

and, for $j = 1, ..., N$,

$$C_j(u) = \exp\left(iuW^\top \mu_j - \frac{u^2}{2}(W^\top \Omega_j W)\right)\left[\left(iW^\top \mu_j - uW^\top \Omega_j W\right)^2 - (W^\top \Omega_j W)\right].$$

Consequently,

$$Var_t[r_{p,t:t+h}] = \frac{\phi^{(2)}_{t:t+h}(0)}{i^2} = \frac{1}{i^2} \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \left\{i^2(W^\top \mu_j - \bar{E})^2 + i^2(W^\top \Omega_j W)\right\}$$

$$+\frac{2}{i^2} \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \left\{i(W^\top \mu_j - \bar{E})(e^\top \frac{d\bar{A}(u)^{h-1}}{du} \mid_{u=0} e_j)\right\}$$

$$+\frac{1}{i^2} \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \left\{e^\top \left(\frac{d^2 \bar{A}(u)^{h-1}}{(du)^2} \mid_{u=0}\right) e_j\right\}$$

$$= e^\top Diag\left((W^\top \mu_1 - \bar{E})^2 + (W^\top \Omega_1 W), ..., (W^\top \mu_N - \bar{E})^2 + (W^\top \Omega_N W)\right) \Pi_{s_t}$$

$$+2W^\top \mu \left(\sum_{l=0}^{h-2} P^{l+1}\right) Diag\left(W^\top \mu_1 - \bar{E}, ..., W^\top \mu_N - \bar{E}\right) \Pi_{s_t}$$

$$+\frac{1}{i^2} \sum_{j=1}^{N} \mathsf{P}(s_t = j \mid I_t) \left\{e^\top \left(\frac{d^2 \bar{A}(u)^{h-1}}{(du)^2} \mid_{u=0}\right) e_j\right\}$$

Observe that

$$\frac{d^2\bar{A}(u)^{h-1}}{(du)^2}\mid_{u=0}=\sum_{l=0}^{h-2}\frac{d\bar{A}(u)^{h-2-l}}{du}\mid_{u=0}B(0)\bar{A}(0)^l$$

$$+\sum_{l=0}^{h-2}\bar{A}(0)^{h-2-l}C(0)\bar{A}(0)^l$$

$$+\sum_{l=0}^{h-2}\bar{A}(0)^{h-2-l}B(0)\frac{d\bar{A}(u)^l}{du}\mid_{u=0}$$

$$=i^2\sum_{l=0}^{h-2}[\sum_{k=0}^{h-3-l}P^{h-3-l-k}\bar{B}(0)P^k]\bar{B}(0)P^l$$

$$+i^2\sum_{l=0}^{h-2}P^{h-2-l}\bar{C}(0)P^l$$

$$+i^2\sum_{l=0}^{h-2}P^{h-2-l}\bar{B}(0)(\sum_{f=0}^{l-1}P^{l-1-f}\bar{B}(0)P^f)$$

where, $\forall u\in\mathbb{R}$,

$$\bar{C}(0)=Diag((W^\top\mu_1)^2+W^\top\Omega_1W,\ldots.,(W^\top\mu_N)^2+W^\top\Omega_NW)P,$$

Thus,

$$Var_t[r_{p,t:t+h}]=e^\top Diag\left((W^\top\mu_1-\bar{E})^2+(W^\top\Omega_1W),\ldots,(W^\top\mu_N-\bar{E})^2+(W^\top\Omega_NW)\right)\Pi_{s_t}$$

$$+2W^\top\mu\left(\sum_{l=1}^{h-1}P^l\right)Diag\left(W^\top\mu_1-\bar{E},\ldots,W^\top\mu_N-\bar{E}\right)\Pi_{s_t}$$

$$+W^\top\mu\left(\sum_{l=1}^{h-2}\sum_{k=1}^{h-l-1}P^k\bar{B}(0)P^{l-1}\right)\Pi_{s_t}+e^\top\bar{C}(0)\left(\sum_{l=1}^{h-1}P^{l-1}\right)\Pi_{s_t}$$

$$+W^\top\mu\left(\sum_{l=1}^{h-2}\sum_{f=1}^{l}P^{l-f+1}\bar{B}(0)P^{f-1}\right)\Pi_{s_t},h\geq 3.$$

which can be written in the following form

$$Var_t[r_{t:t+h}] = Var[r_{t:t+h} \mid I_t] = \left(\Pi_{s_t}^{\top} \otimes I_n\right) \bar{\Omega}_t + 2\mu \left(\sum_{l=1}^{h-1} P^l\right) Diag(\Pi_{s_t}) \left[\mu - e^{\top} \otimes \bar{\mu}_t\right]^{\top}$$

$$+2\mu \left[\sum_{l=1}^{h-2} \sum_{k=1}^{h-l-1} P^k Diag(P^l \Pi_{s_t})\right] \mu^{\top}$$

$$+ \left(\Pi_{s_t}^{\top} \otimes I_n\right) \left(\left(\sum_{l=1}^{h-1} P^l\right)^{\top} \otimes I_n\right) \Omega,\ h \geq 3.$$

■

Proof of Proposition 12. Proposition 12 can be deduced from Proposition 11 using the law of iterated expectations. ■

Appendix 2: Existence and Uniqueness of the solution of Equation (3.15)

Proof of the existence of a solution of Equation (3.15). We have to show that the following function:

$$f(VaR^\alpha) = -\pi[\mathsf{P}_t(r_{t+1} < -VaR^\alpha) - \alpha] = 0. \tag{3.41}$$

has a solution. To do so we need to check whether the function (3.41) satisfies the following two conditions:

1. $f(VaR^\alpha)$ is monotone,

2. there exists some x_1 and x_2 such that:

$$f(x_1) < (>)0 \text{ and } f(x_2) > (<)0.$$

The first condition follows from one of the properties of the probability distribution function. We know that $\mathsf{P}_t(r_{t+1} < -VaR^\alpha)$ is monotonically increasing, then $f(VaR^\alpha)$ is monotonically decreasing because of the factor $-\pi < 0$. The second condition can derived from other properties of the probability distribution function. For $x \in \mathbb{R}$, we have

$$\lim_{x \to -\infty} \mathsf{P}_t(r_{t+1} < x) = 0,$$

$\Longrightarrow$

$$\lim_{x \to -\infty} -\pi[\mathsf{P}_t(r_{t+1} < x) - \alpha] = \alpha\pi > 0, \text{ for } 0 < \alpha < 1.$$

Similarly,

$$\lim_{x \to +\infty} \mathsf{P}_t(r_{t+1} < x) = 1.$$

$\Longrightarrow$

$$\lim_{x \to +\infty} -\pi[\mathsf{P}_t(r_{t+1} < x) - \alpha] = -\pi(1-\alpha) < 0, \text{ for } 0 < \alpha < 1.$$

Thus, $f(VaR^{\alpha})$ satisfies the above two conditions and admits a solution. ■

Proof of the uniqueness of the solution of Equation (3.15). The uniqueness of the solution to equation (3.41) is immediate because $\mathsf{P}_t(r_{t+1} < -VaR^{\alpha})$ is a strictly increasing function and $f(VaR^{\alpha})$ is a strictly decreasing function. ■

Appendix 3: Empirical Results

Table 4: Summary statistics for S&P 500 index returns, 1988-1999.

	Mean	*St.Dev.*	*Median*	*Skewness*	*Kurtosis*
Daily returns	0.0650	0.8653	0.0458	−0.4875	9.2644

Note: This table summarizes the daily return distributions for the S&P 500 index. The sample covers the period from January 1988 to May 1999 for a total of 2959 trading days.

Table 5: Summary statistics for TSE 300 index returns, 1988-1999.

	Mean	*St.Dev.*	*Median*	*Skewness*	*Kurtosis*
Daily returns	0.0365	0.6752	0.0415	−0.9294	12.1580

Note: This table summarizes the daily return distributions for the TSE 300 index. The sample covers the period from January 1988 to May 1999 for a total of 2959 trading days.

Table 6: Parameter estimates for the bivariate Markov switching model.

Parameters	*Value*	*St.Error*	*T-Statistics*
p_{11}	0.95535	0.00876073	109.05
p_{12}	0.17844	0.032071	5.56384
μ_{11}	0.08903	0.0141436	6.29475
μ_{21}	0.073807	0.0103029	7.1637
μ_{12}	−0.032714	0.0452048	−0.723676
μ_{22}	−0.11184	0.0496294	−2.25349
$\sigma^2_{11,1}$	0.40985	0.018317	22.3757
$\sigma^2_{22,1}$	0.20396	0.00911366	22.3798
$\sigma_{21,1}$	0.16146	0.0099733	16.1893
$\sigma^2_{11,2}$	2.0895	0.162652	12.8465
$\sigma^2_{22,2}$	1.4354	0.124571	11.5231
$\sigma_{21,2}$	1.2653	0.119861	10.5562

Note: This table shows the estimation results for the two-state bivariate Markov switching model. The second column represents the parameter estimation results for the elements of the transition probability matrix, mean returns in states 1 and 2, and the variance-covariance matrix in states 1 and 2, respectively. The third column represents the standard errors of the estimates. Finally, the fourth column gives the t-statistics.

Figure 1: S&P 500, Daily returns, 1988-1999

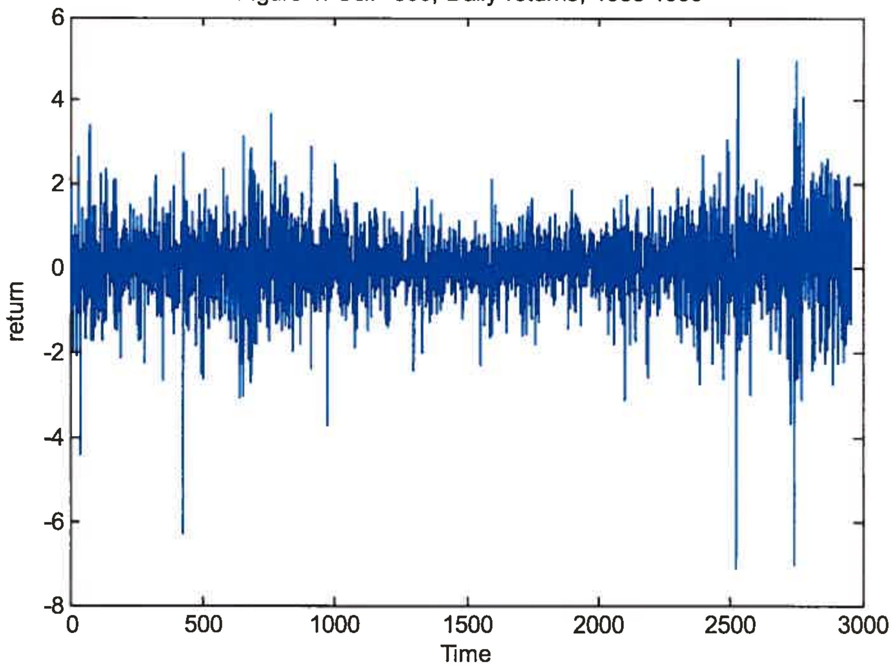


Figure 2: TSE 300, Daily returns, 1988-1999

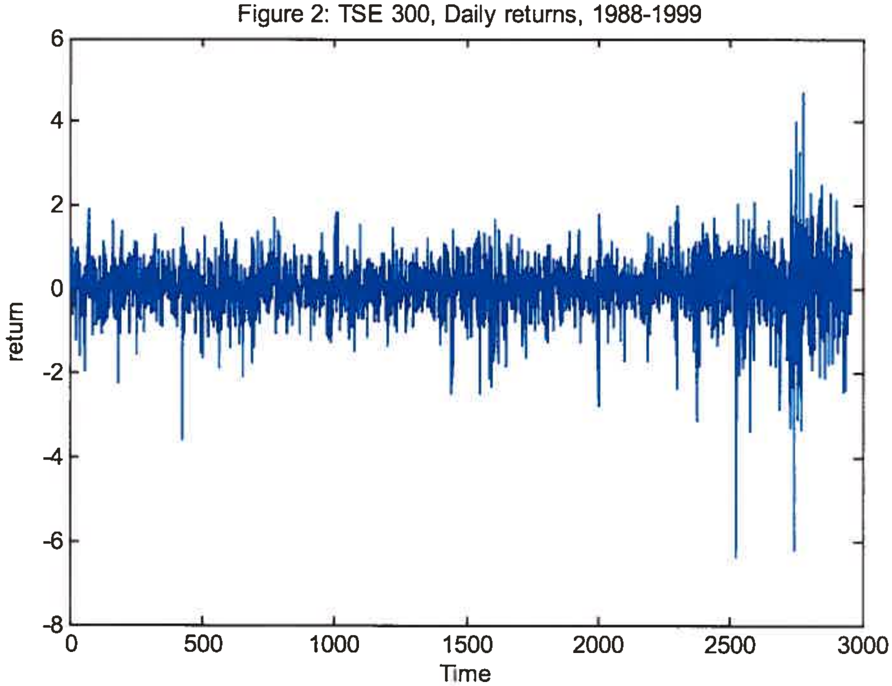


Figure 3: Filtred probabilities of regimes 1 and 2

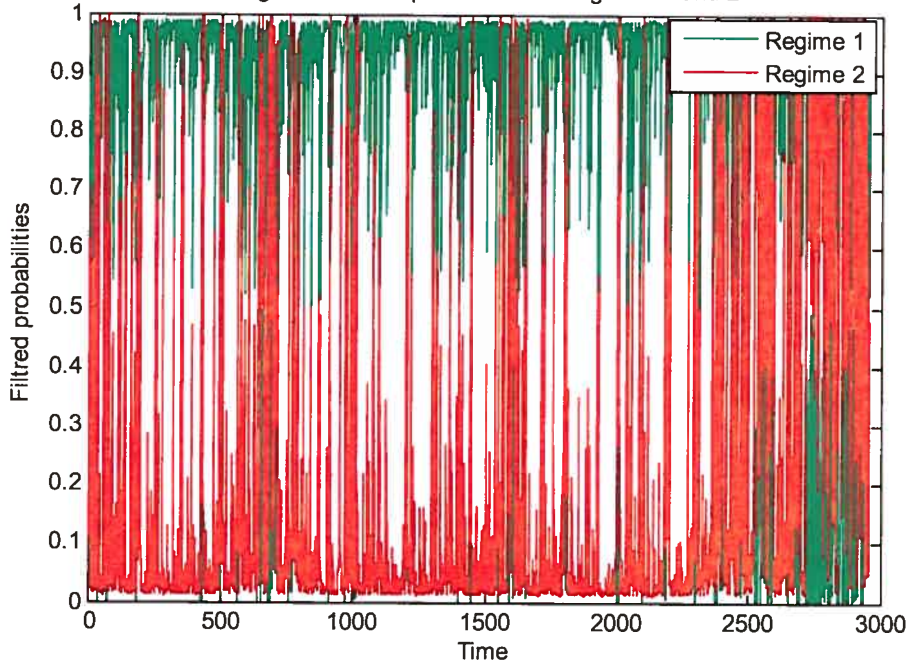


Figure 4: Smoothed probabilities of regimes 1 and 2

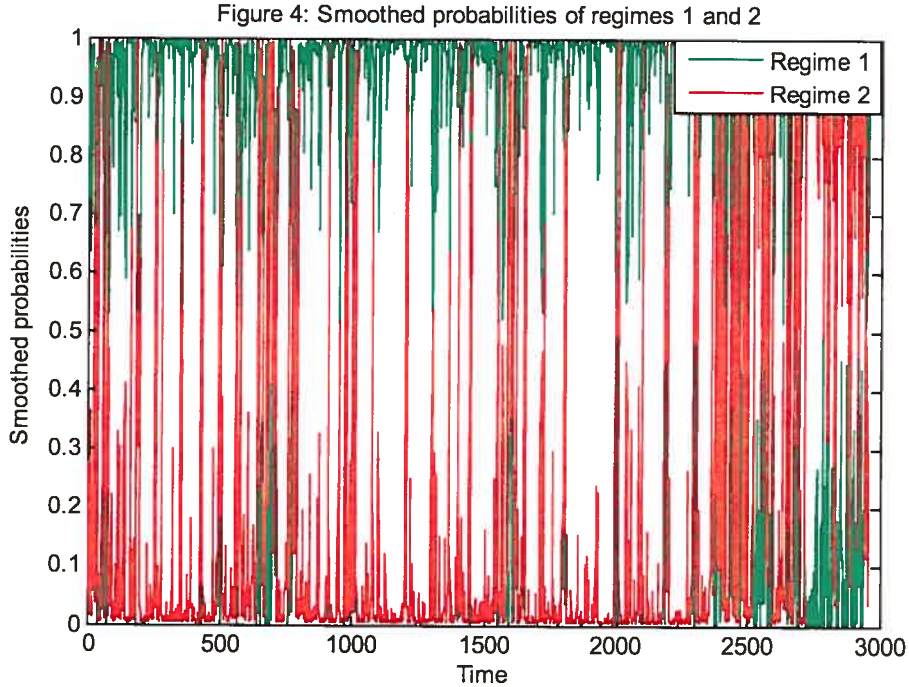


Figure 5: One period ahead variance

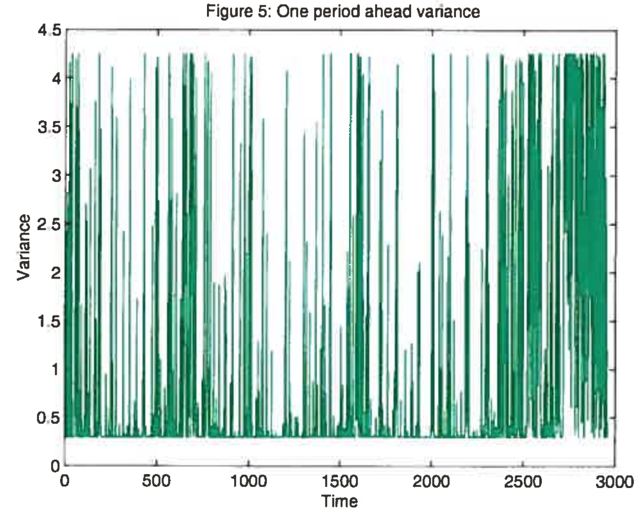


Figure 6: 5 periods ahead variance

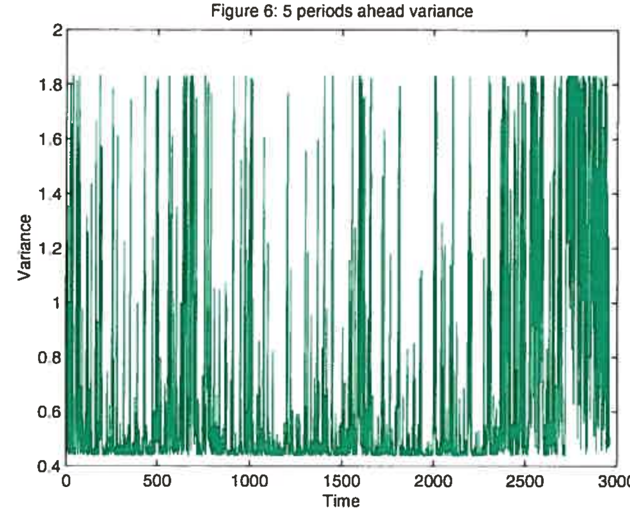


Figure 7: 15 periods ahead variance

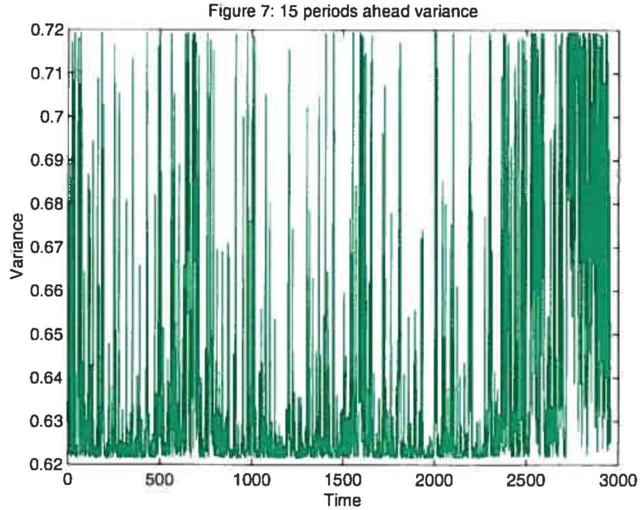


Figure 8: h periods ahead variance of the portfolio's return

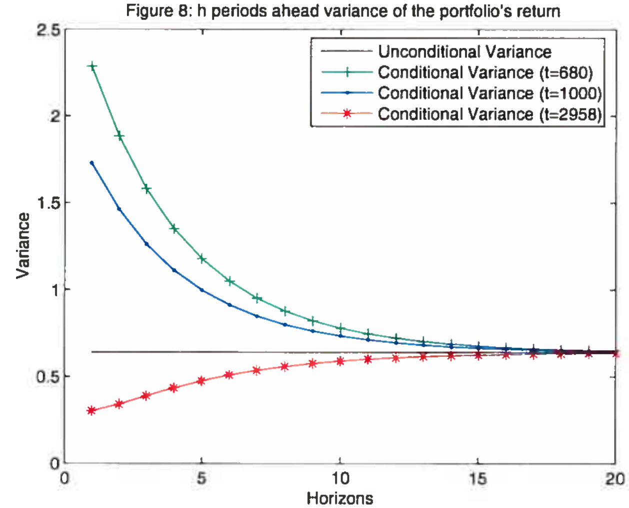


Figure 9: One period ahead 5% VaR

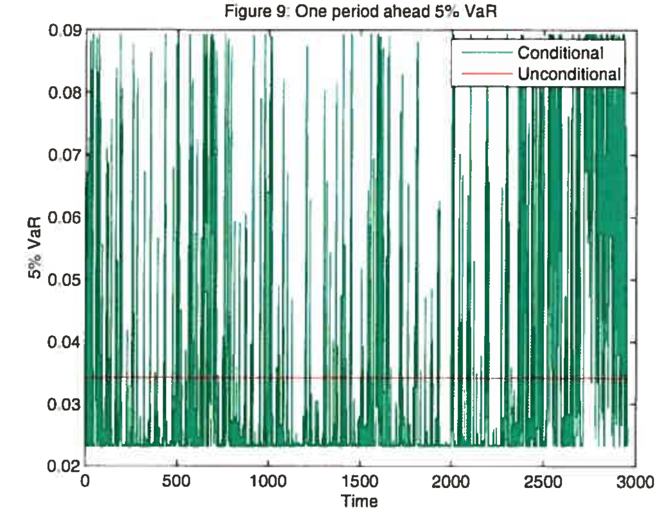


Figure 10: 5 periods ahead 5% VaR

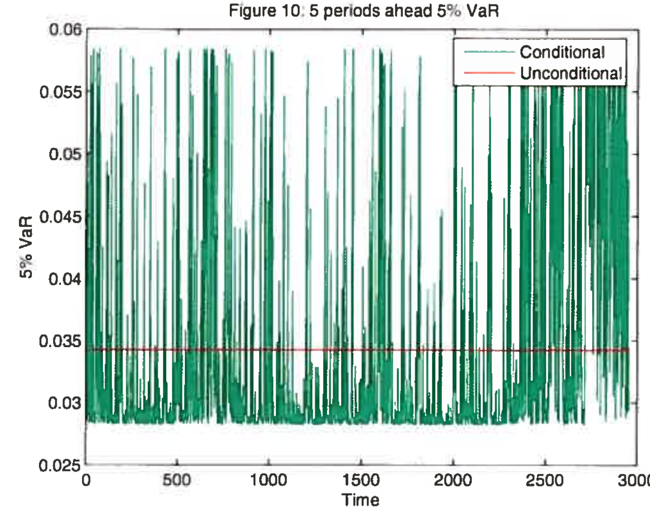


Figure 11: 15 periods ahead 5% VaR

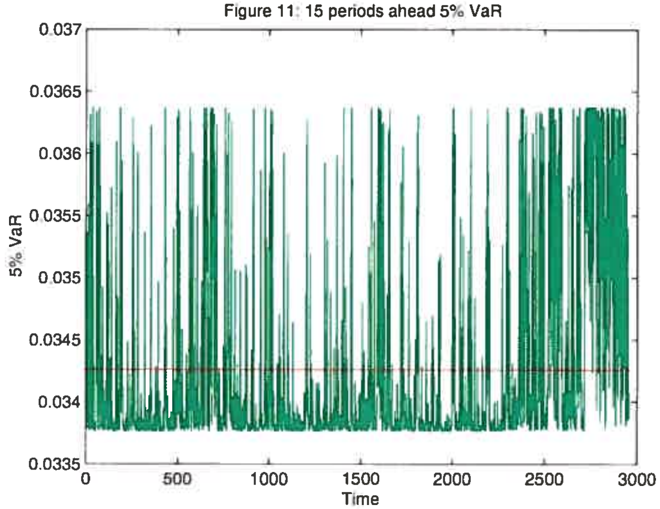


Figure 12: h periods ahead 5% VaR

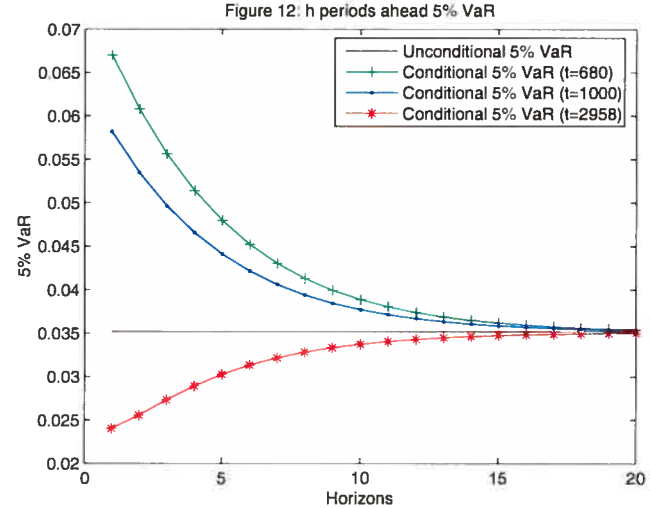


Figure 13: h periods ahead 10% VaR

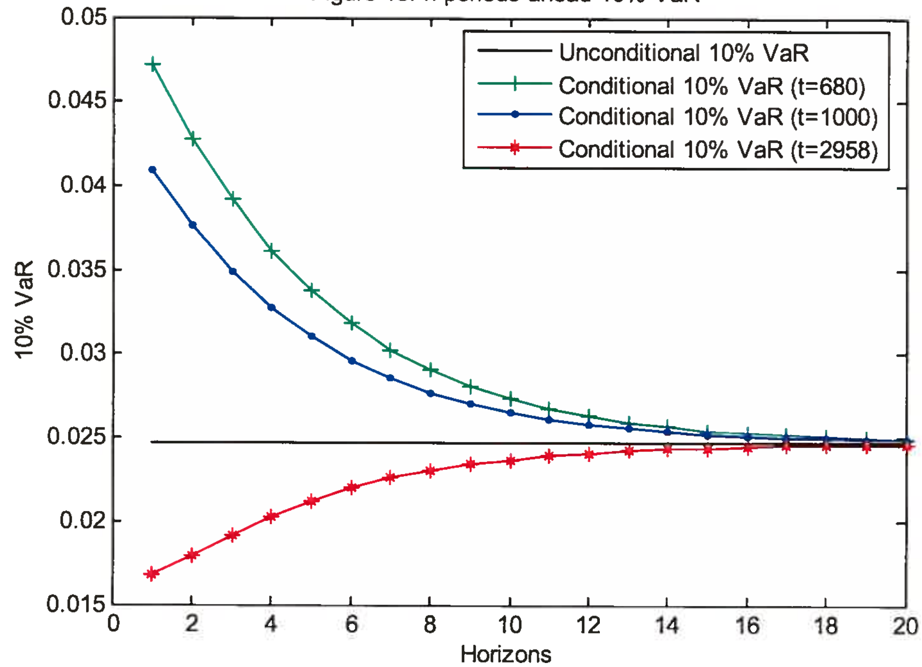


Figure 14: Unconditional h periods ahead simple Mean-Variance Efficient Frontier

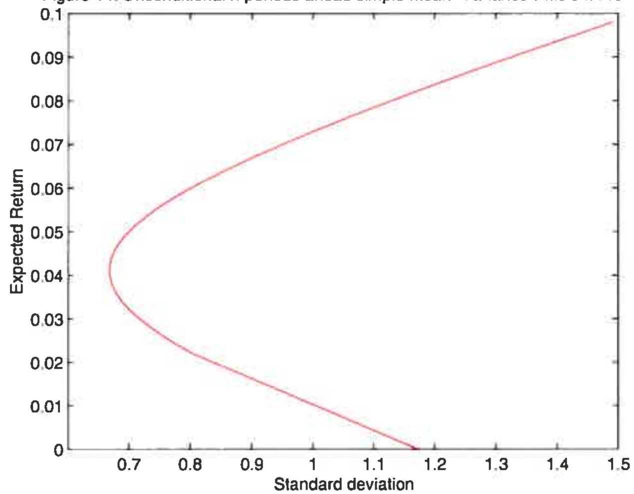


Figure 15: h periods ahead simple Mean-Variance Efficient Frontier (t=2958)

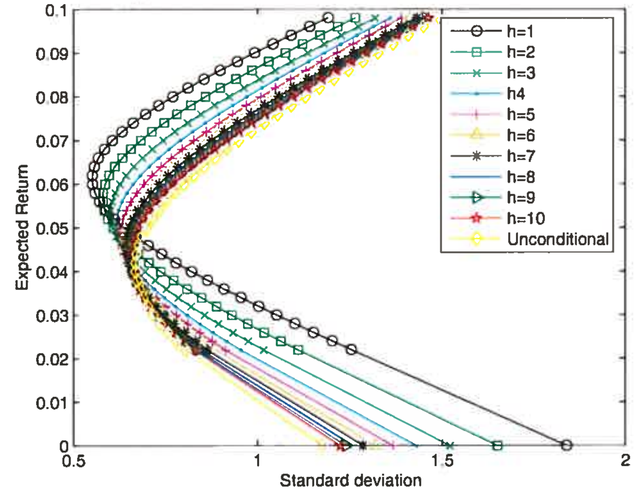


Figure 16: h periods ahead simple Mean-Variance Efficient Frontier (t=680)

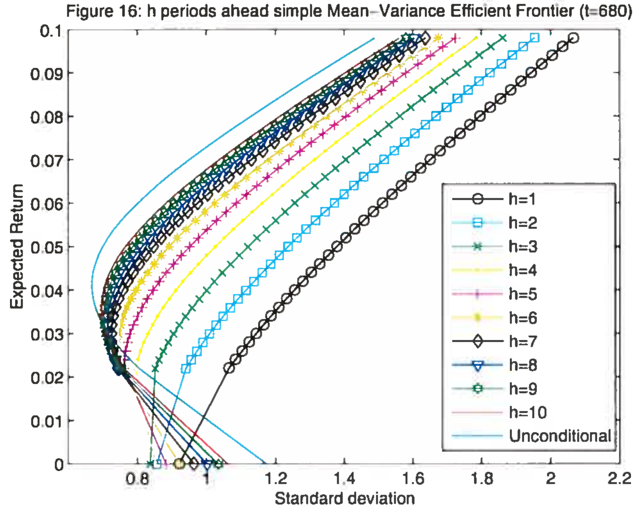


Figure 18: h periods ahead simple Mean-Variance Efficient Frontier (t=1000)

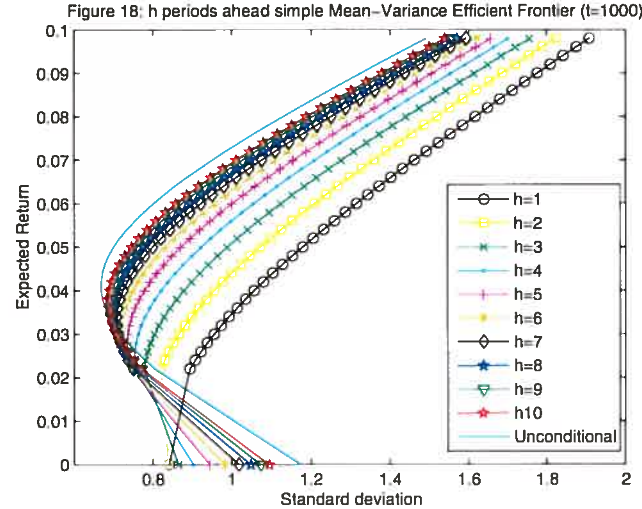


Figure 18: One period ahead Sharpe Ratio

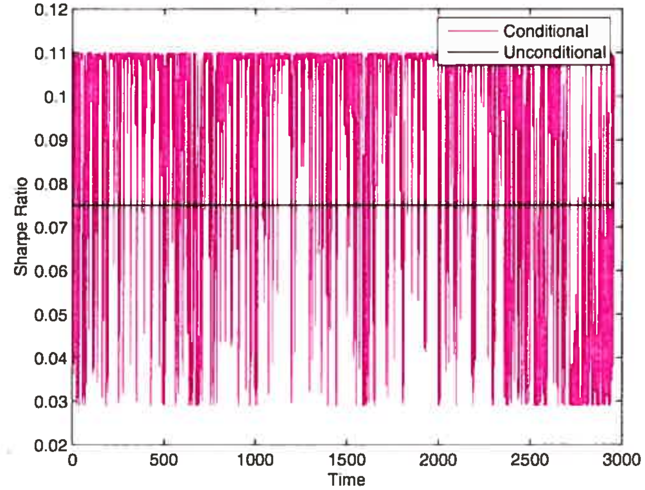


Figure 19: 5 periods ahead Sharpe Ratio

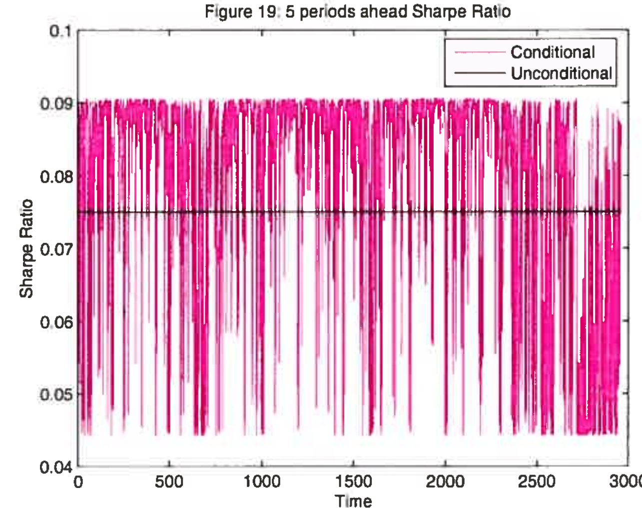


Figure 20: 15 periods ahead Sharpe Ratio

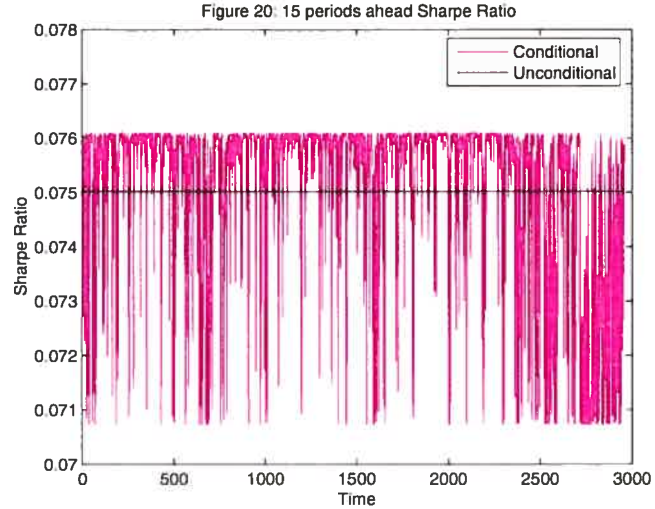


Figure 21: h periods ahead Sharpe Ratio

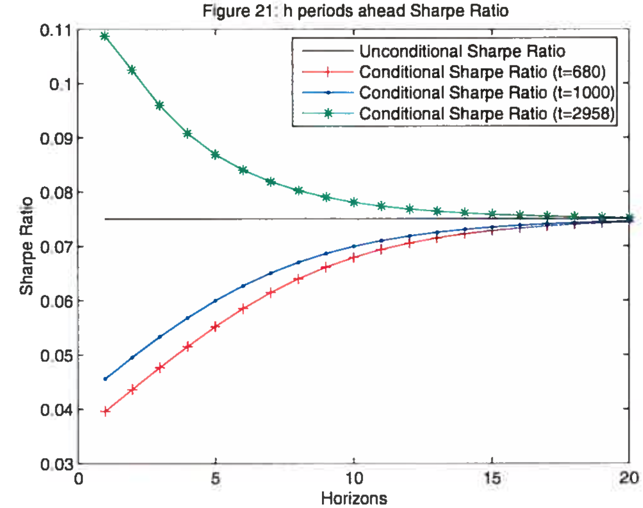


Figure 22: One period ahead aggregated variance

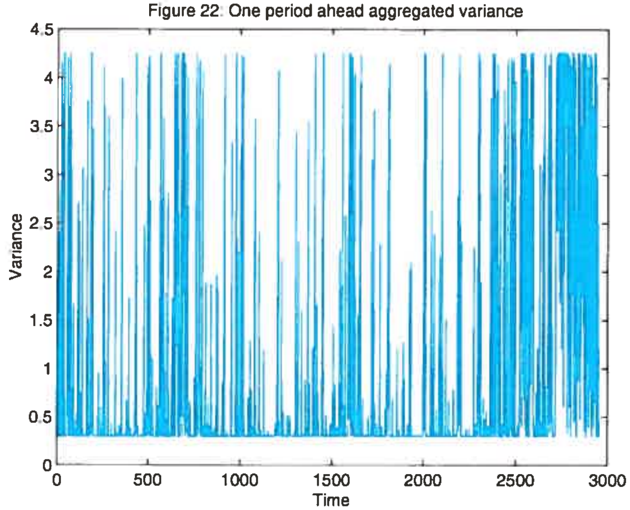


Figure 23: 5 periods ahead aggregated variance

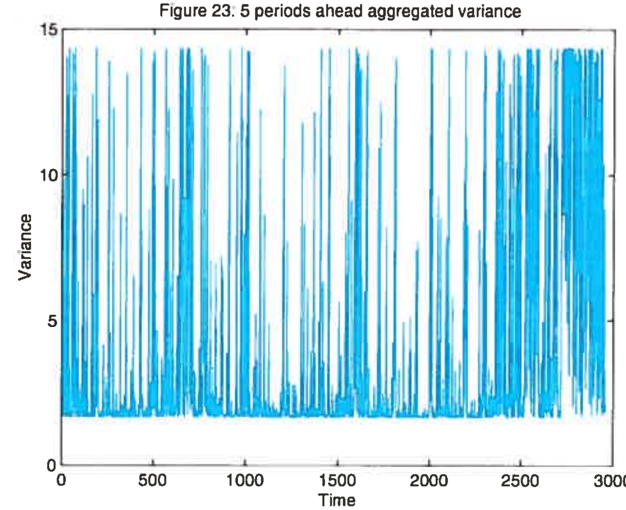


Figure 24: 15 periods ahead aggregated variance

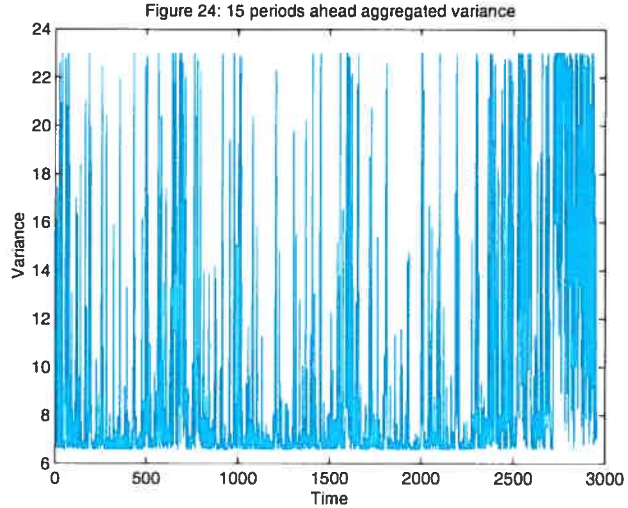


Figure 25: h periods ahead variance of the aggregated portfolio return

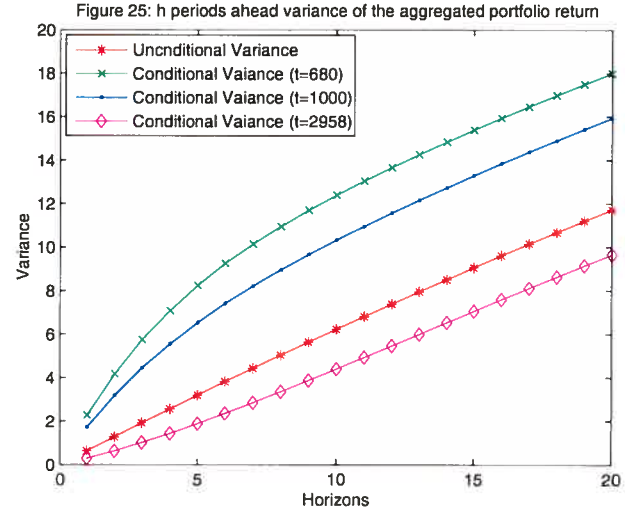


Figure 26: One period ahead aggregated 5% VaR

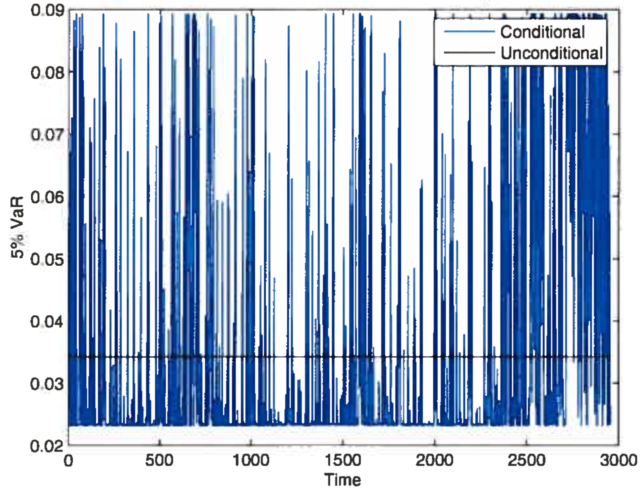


Figure 27: 5 periods ahead aggregated 5% VaR

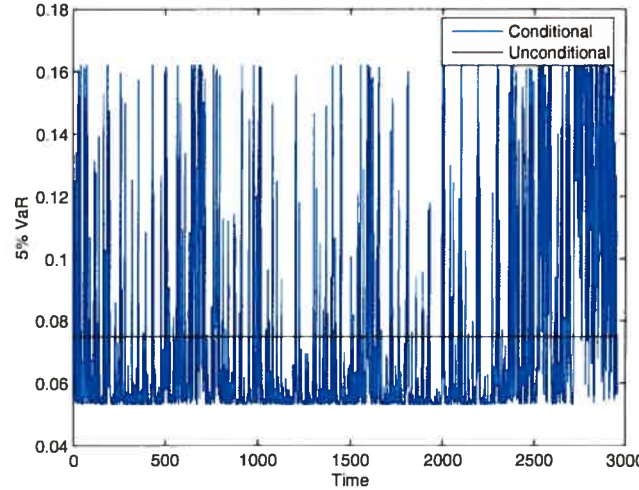


Figure 28: 15 periods ahead aggregated 5% VaR

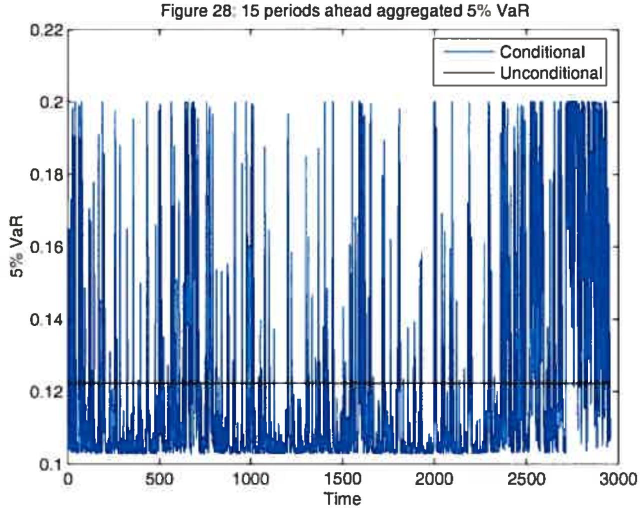


Figure 29: h periods ahead aggregated 5% VaR

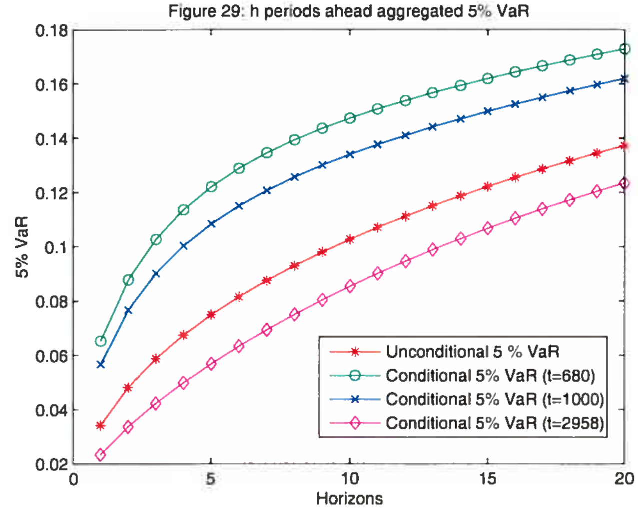
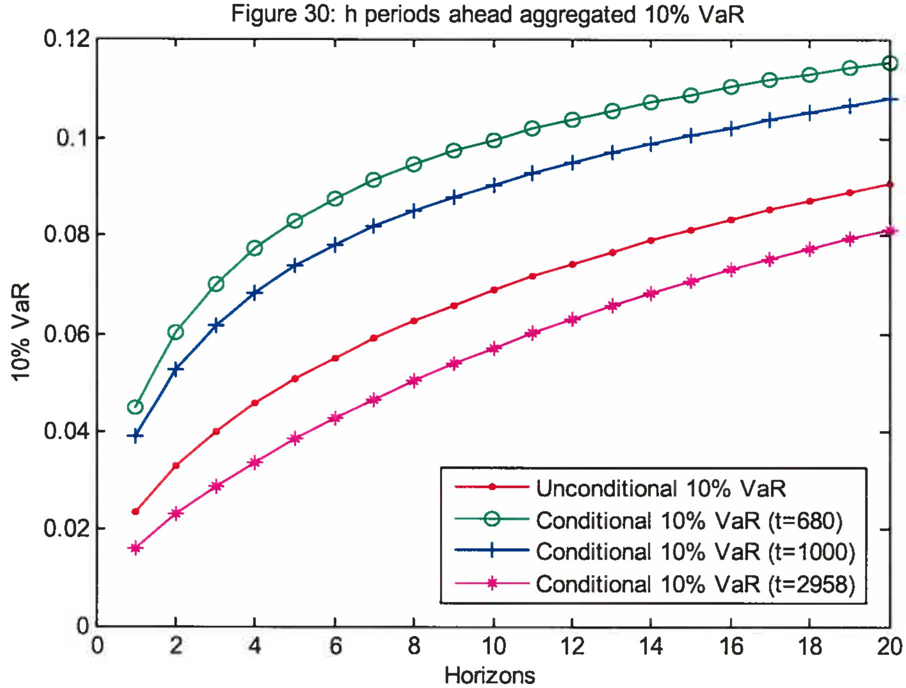


Figure 30: h periods ahead aggregated 10% VaR



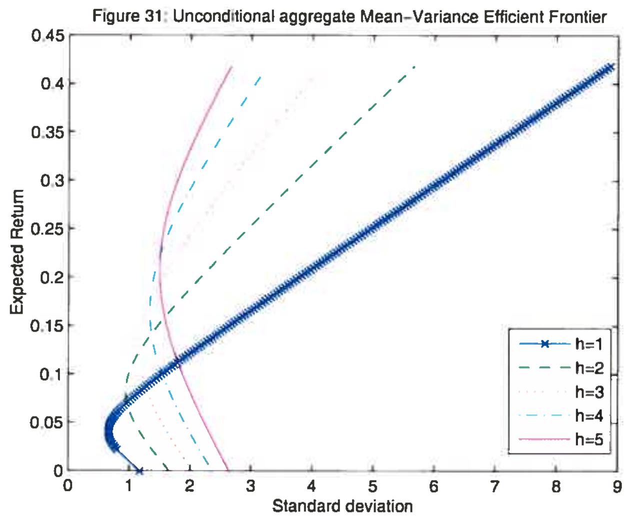


Figure 31: Unconditional aggregate Mean-Variance Efficient Frontier
Expected Return
Standard deviation
h=1
h=2
h=3
h=4
h=5

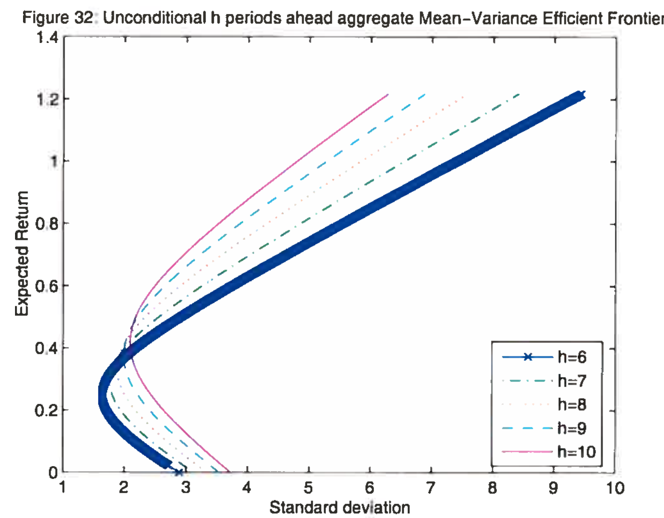


Figure 32: Unconditional h periods ahead aggregate Mean-Variance Efficient Frontier
Expected Return
Standard deviation
h=6
h=7
h=8
h=9
h=10

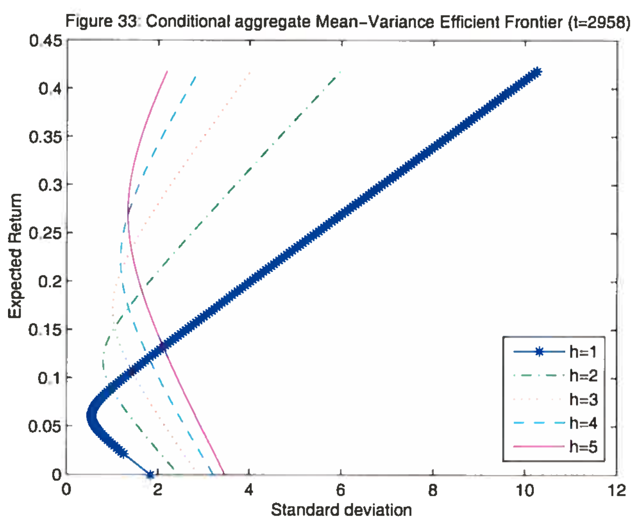


Figure 33: Conditional aggregate Mean-Variance Efficient Frontier (t=2958)
Expected Return
Standard deviation
h=1
h=2
h=3
h=4
h=5

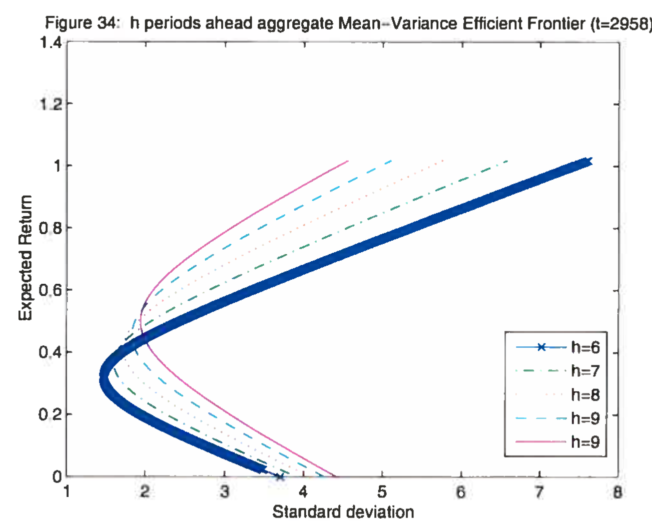


Figure 34: h periods ahead aggregate Mean-Variance Efficient Frontier (t=2958)
Expected Return
Standard deviation
h=6
h=7
h=8
h=9
h=9

Figure 35: One period ahead aggregated Sharpe Ratio

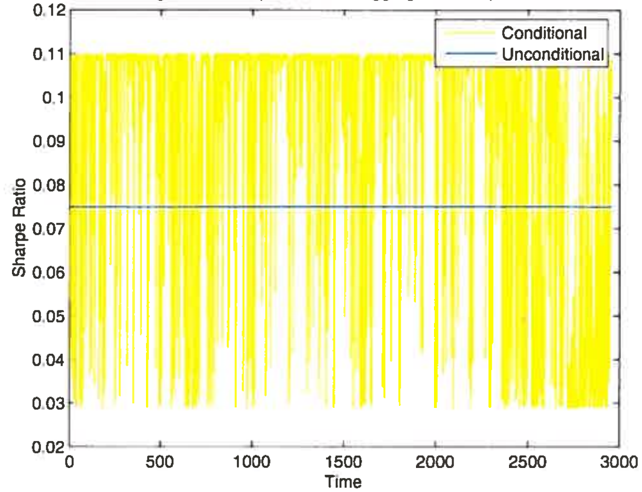


Figure 36: 5 periods ahead aggregated Sharpe Ratio

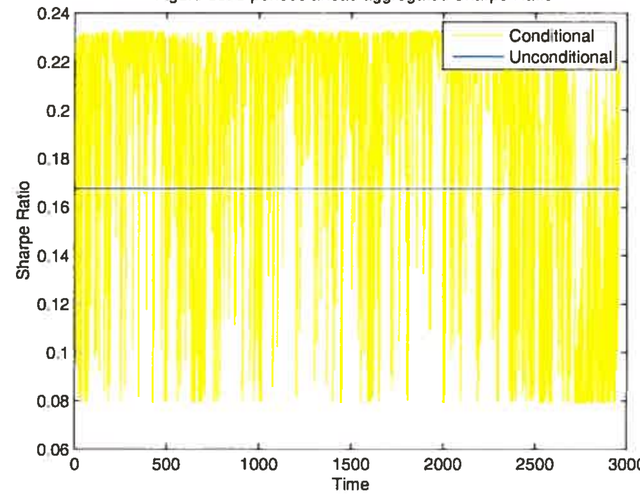


Figure 37: 15 periods ahead aggregated Sharpe Ratio

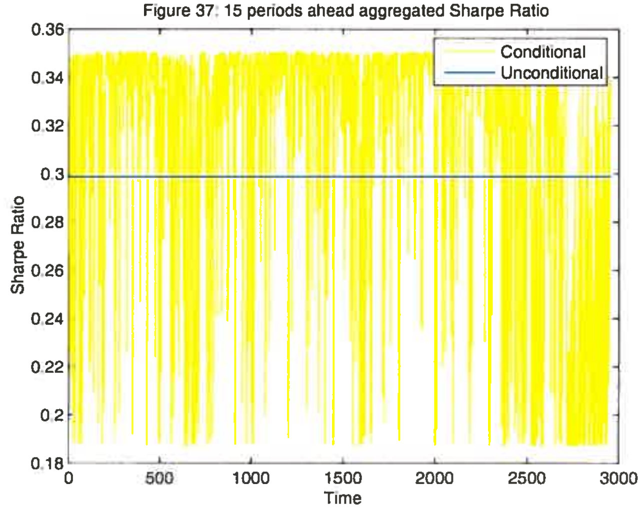
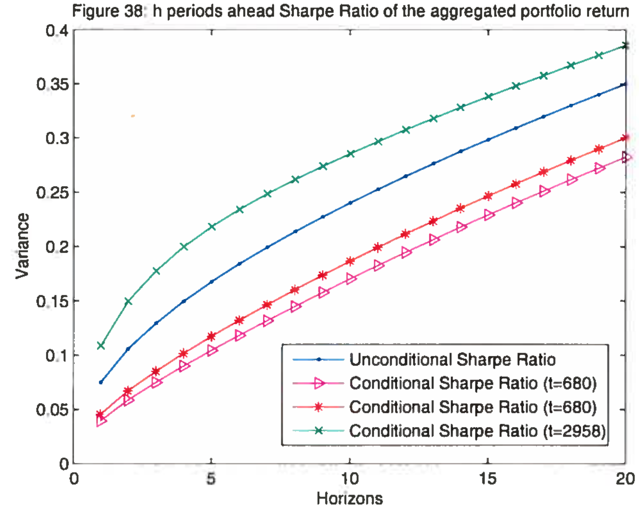


Figure 38: h periods ahead Sharpe Ratio of the aggregated portfolio return



Chapter 4

Exact optimal and adaptive inference in linear and nonlinear models under heteroskedasticity and non-normality of unknown forms

4.1 Introduction

In practice, most economic data are heteroskedastic and non-normal. In the presence of some types of heteroskedasticity, the parametric tests proposed to improve inference may exhibit poor size control and/or low power. For example, when there is a break in the disturbance variance, our simulation results show that the usual test statistic based on White's (1980) correction of the variance, which is supposed to be robust against heteroskedasticity, has very poor power. Other forms of heteroskedasticity for which the usual tests are less powerful are exponential variance and $GARCH$ with one or several outliers.[1] At the same time, many *exact* parametric tests developed in the literature typically assume normal disturbances. The latter assumption is unrealistic and, in the presence of heavy tails or asymmetric distributions, our simulation results show that these tests may not perform very well in terms of power and do not control size. Furthermore, the statistical procedures developed for inference on parameters of *nonlinear* models are typically based on asymptotic approximations and there are only a few exact inference methods outside the linear model framework. However, these approximations may be invalid, even in large samples [see Dufour (1997)]. The present chapter aims to propose exact tests which work under more realistic assumptions. We derive simple optimal sign-based tests to test the values of parameters in linear and nonlinear regression models. These tests are valid under weak distributional assumptions such as heteroscedasticity of unknown form and non-normality.

Several authors have provided theoretical arguments for why the existing parametric tests about the mean of i.i.d. observations fail under weak distributional assumptions, such as non-normality and heteroskedasticity of unknown form. Bahadur and Savage

[1] One characteristic of the financial markets is the presence of episodic occasional of crashes and rallies, as shown by the extreme values in Figure 1 [see appendix], which represents a time series plot of daily returns of the S&P 500 stock price index. These extreme values can be viewed as introducing outliers in the $GARCH$ model. Moreover, it may occur that financial returns series contain other atypical observations such as additive or innovation outliers. The reader can consult Hotta and Tsay (1998) for a recent classification of outliers in $GARCH$ models and Friedman and Laibson (1989) for the economic arguments for the possible presence of atypical observations.

(1956) show that under weak distributional assumptions on the error terms, it is not possible to obtain a valid test for the mean of i.i.d. observations even for large samples. Many other hypotheses about various moments of i.i.d. observations lead to similar difficulties. This can be explained by the fact that moments are not empirically meaningful in non-parametric models or models with weak assumptions. Lehmann and Stein (1949) and Pratt and Gibbons (1981, sec. 5.10) show that conditional sign methods were the only possible way of producing valid inference finite sample procedures under conditions of heteroskedasticity of unknown form and non-normality. More discussion about the statistical inference problems in non-parametric models can be find in Dufour (2003).

This chapter introduces a new sign-based tests in the context of linear and nonlinear regression models. The proposed tests are exact, distribution-free, robust against heteroskedasticity of unknown form, and they may be inverted to obtain confidence regions for the vector of unknown parameters. These tests are derived under assumptions that the disturbances in regression models are independent, but not necessarily identically distributed, with a null median conditional on the explanatory variables. A few sign-based test procedures have been developed in the literature. In the presence of only one explanatory variable, Campbell and Dufour (1995, 1997) propose nonparametric analogues of the *t-test*, based on sign and signed rank statistics, that are applicable to a specific class of feedback models including both Mankiw and Shapiro's (1986) model and the random walk model. These tests are exact even if the disturbances are asymmetric, non-normal, and heteroskedastic. Boldin, Simonova and Tyurin (1997) propose locally optimal sign-based inference and estimation for linear models. Coudin and Dufour (2005) extend the work by Boldin and al. (1997) to some forms of statistical dependence in the data. Wright (2000) proposes variance-ratio tests based on the ranks and signs to test the null hypothesis that the series of interest is a martingale difference sequence.

The present chapter address the issue of the optimality and seeks to derive point-optimal tests based on sign statistics. Point-optimal tests are useful in a number of ways and they are most attractive for problems in which the size of the parameter space can be restricted by theoretical considerations. Because of their power properties, these tests are particularly attractive when testing one economic theory against another, for example a new theory against an existing theory. They would ensure optimal power at given point and, depending on the structure of the problem, yield good power over the entire parameter space. Another interesting feature is that they can be used to trace out the maximum attainable power envelope for a given testing problem. This power envelope provides an obvious benchmark against which test procedures can be evaluated. More discussion about the usefulness of point-optimal tests can be found in King (1988). Many papers have derived point-optimal tests to improve inference in the context of some economic problems. Dufour and King (1991) use point-optimal tests to do inference on the autocorrelation coefficient of a linear regression model with first-order autoregressive normal disturbances. Elliott, Rothenberg, and Stock (1996) derive the asymptotic power envelope for point-optimal tests of a unit root in the autoregressive representation of a Gaussian time series under various trend specifications. Recently, Jansson (2005) derives an asymptotic Gaussian power envelope for tests of the null hypothesis of cointegration and proposes a feasible point-optimal cointegration test whose local asymptotic power function is found to be close to the asymptotic Gaussian power envelope.

Since the point-optimal conditional sign tests depend on the alternative hypothesis, we propose an adaptive approach based on split-sample technique to choose an alternative that makes the power curve of the point-optimal conditional sign test close to that of the power envelope.[2] The idea is to divide the sample into two independent parts and to use the first one to estimate the value of the alternative and the second one to compute the point-optimal conditional sign test statistic. The simulation results show that using approximately 10% of sample to estimate the alternative yields a power which is typically very close to the power envelope. We present a Monte Carlo study to assess the performance of the proposed "quasi"-point-optimal conditional sign test by compar-

[2]For more details about the sample-split technique, the reader can consult Dufour and Torrès (1998) and Dufour and Jasiak (2001).

ing its size and power to those of some common tests which are supposed to be robust against heteroskedasticity. The results show that our procedure is superior.

The plan of this chapter is as follows. In section 4.2, we present the general framework that we need to derive the point-optimal conditional sign tests (hereafter POS tests or $POST$). In section 4.3, we derive POS tests to test the value of parameters in the context of linear and nonlinear regression models. In section 4.4, we study the power properties of the POS test and we propose an adaptive approach to choose the optimal alternative. In section 4.5, we discuss the construction of the point-optimal sign confidence region (hereafter $POSC$) using projection techniques. In section 4.6, we present a Monte Carlo simulation to assess the performance of the POS test by comparing its size and power to those of some popular tests. The conclusion relating to the results is given in section 4.7. Technical proof are given in section 4.8.

4.2 Framework

In this section, we introduce a framework for deriving point-optimal conditional sign tests in the context of some statistical problems such as testing the parameters in linear and nonlinear regression models. Point-optimal tests are useful in a number of ways and they are most attractive for problems in which the size of the parameter space can be restricted by theoretical considerations. They would ensure optimal power at given point and, depending on the structure of the problem, yield good power over the entire parameter space. In our development we consider simple hypotheses that can be constant or not. We use Neyman-Pearson lemma to derive conditional sign-based tests, for both hypotheses.

In the remainder of the chapter we suppose that $\{y_t\}_{t=1}^{n}$ is a random sample and, for $t = 1, \ldots, n$,

$$y_t \text{ are independent.} \tag{4.1}$$

We define the following vector of signs

$$U(n) = [s(y_1), ..., s(y_n)]',$$

where, for $t = 1, ..., n$,

$$s(y_t) = \begin{cases} 1, \ if \ y_t \geq 0, \\ \\ 0, \ if \ y_t < 0. \end{cases}$$

Here we assume that there is no probability mass at zero, or, for $t = 1, ..., n$, $\mathsf{P}[y_t = 0]$. This holds, for example, when y_t is a continuos variable.

4.2.1 Point-optimal sign test for a constant hypothesis

Let $y = (y_1, ..., y_n)'$ be an observable $n \times 1$ vector of independent random variables such that $\mathsf{P}[y_t \geq 0] = p$. We wish to test:

$$\bar{H}_0 : p = \delta \ , \ \delta \in \Omega \tag{4.2}$$

against

$$\bar{H}_1 : p = \eta, \ \eta \in \Theta$$

where Ω and Θ are subsets of $[0, 1]$. $\bar{H}_0$ and $\bar{H}_1$ correspond to composite hypotheses and represent very general testing problems. If we consider the following test problem which consists in testing

$$H_0 : p = p_0 \tag{4.3}$$

against

$$H_1 : p = p_1 \tag{4.4}$$

where p_0 and p_1 are fixed and known, then we have simple null and alternative hypothesis. Here we consider an optimal test in the Neyman-Pearson sense which minimizes the Type

II error, or maximize the power, under the constraint

$$\mathsf{P}[\text{reject } H_0 \mid H_0] \leq \alpha.$$

If we denote the density of y under the null by $f(y \mid H_0)$ and its density under the alternative by $f(y \mid H_1)$, then the Neyman-Pearson lemma [see e.g. Lehmann (1959, p.65)] implies that rejecting H_0 for large values of

$$s = \frac{f(y \mid H_1)}{f(y \mid H_0)} \tag{4.5}$$

is the most powerful test. In this case the critical value, denoted c, for the test statistic is given by the smallest constant c such that

$$\mathsf{P}[s > c \mid H_0] \leq \alpha$$

where α is the desired level of significance or a Type I error. The choice of a significance level α is usually somewhat arbitrary, since in most situations there is no precise limit to the probability of a Type I error that can be tolerated. Standard values, such as 0.01 or 0.05, were originally chosen to effect a reduction in the tables needed for carrying out various test. However, the choice of significance level should take into consideration the power that the test will achieve against the alternative of interest. Rules for choosing α in relation to the attainable power are discussed by Lehmann (1958), Arrow (1960), Sanathanan (1974), and Lehmann and Romano (2005).

For our statistical problem which consists to test for some values of $\mathsf{P}[y_t \geq 0]$, the likelihood function of the sample $\{y_t\}_{t=1}^{n}$ is given by:

$$L(U(n), p) = \prod_{t=1}^{n} \mathsf{P}[y_t \geq 0]^{s(y_t)}(1 - \mathsf{P}[y_t \geq 0])^{1-s(y_t)}. \tag{4.6}$$

The Neyman-Pearson test is based on the values of likelihood function under H_0 and H_1.

Under H_0, the function (4.6) has the form

$$L_0(U(n), p_0) = \prod_{t=1}^{n} p_0^{s(y_t)}(1-p_0)^{1-s(y_t)} = p_0^{S_n}(1-p_0)^{n-S_n},$$

where $S_n = \sum_{t=1}^{n} s(y_t)$ and, under the alternative H_1, it takes the form

$$L_1(U(n), p_1) = \prod_{t=1}^{n} p_1^{s(y_t)}(1-p_1)^{1-s(y_t)} = p_1^{S_n}(1-p_1)^{n-S_n}.$$

The likelihood ratio is then given by:

$$\frac{L_1(U(n), p_1)}{L_0(U(n), p_0)} = \prod_{t=1}^{n}\left\{\left(\frac{p_1}{p_0}\right)^{s(y_t)}\left(\frac{1-p_1}{1-p_0}\right)^{1-s(y_t)}\right\} = \left(\frac{p_1}{p_0}\right)^{S_n}\left(\frac{1-p_1}{1-p_0}\right)^{n-S_n}. \tag{4.7}$$

For simplicity of exposition we assume that $p_0, p_1 \neq 0, 1$. This allows us to work with the log-likelihood function which simplifies the expression for the test statistic. When $p_0 = 0, 1$, we could work directly with likelihood function. From (4.7) we deduce the log-likelihood function:

$$\ln\left\{\frac{L_1(U(n), p_1)}{L_0(U(n), p_0)}\right\} = S_n\left\{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)\right\} + n\ln\left(\frac{1-p_1}{1-p_0}\right).$$

The best test of H_0 against H_1 based on $s(y_1), ..., s(y_n)$ rejects H_0 when

$$\ln\left\{\frac{L_1(U(n), p_1)}{L_0(U(n), p_0)}\right\} > c. \tag{4.8}$$

If we choose the alternative p_1 such that $p_1 > p_0 > 0$, then the above test is equivalent to rejecting H_0 when

$$S_n > c_1 \equiv \frac{c - n\ln\left(\frac{1-p_1}{1-p_0}\right)}{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)},$$

where c_1 satisfies

$$\mathsf{P}[S_n > c_1 \mid H_0] \leq \alpha.$$

This test is the same for all $p_1 > p_0$. Similarly, if $0 < p_1 < p_0$, the test (4.8) is equivalent to rejecting when

$$S_n < c_1 \equiv \frac{c - n\ln\left(\frac{1-p_1}{1-p_0}\right)}{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)}.$$

where c_1 satisfies

$$\mathsf{P}[S_n < c_1 \mid H_0] \leq \alpha.$$

Thus, under assumption (4.1) and for $p_1 > p_0 > 0$ the test with critical region

$$C = \{(y_1, \ldots, y_n) : S_n > c_1\}$$

is the best point-optimal conditional sign test for the null hypothesis (4.3) against the alternative (4.4). Similarly, for $0 < p_1 < p_0$, the critical region of the best point-optimal sign test is given by

$$C = \{(y_1, \ldots, y_n) : S_n < c_1\}.$$

The value of c_1 is chosen so that

$$\mathsf{P}((y_1, \ldots, y_n) \in C \mid H_0) \leq \alpha.$$

In both cases, i.e. for $p_1 > p_0 > 0$ and $0 < p_1 < p_0$, the test statistic is given by

$$S_n = \sum_{t=1}^{n} s(y_t).$$

Under H_0, S_n follows a binomial distribution $Bi(n, p_0)$, i.e. $\mathsf{P}(S_n = i) = C_n^i p_0^i (1-p_0)^{n-i}$, for $i = 0, 1, \ldots, n$, where $C_n^i = \frac{n!}{[i!(n-i)!]}$. This result corresponds to a *uniformly most powerful* (*UMP*) test, since S_n does not depend on the alternative p_1.

Example 5 (Backtesting Value-at-Risk) *Backtesting Value-at-Risk (VaR) is a key part of the internal model's approach to market risk management as laid out by the Basle Committee on Banking Supervision (1996).*[3] *Christoffersen (1998) proposes a test for*

[3] *For more discussion about the Backtesting VaR, the reader can consult Christoffersen and Pelletier*

unconditional coverage of VaR based on the standard likelihood ratio test.

Consider a time series of daily ex post portfolio returns, R_t, and a corresponding time series of ex ante VaR forecasts, $VaR_t(p)$, with promised coverage rate p, such that $\mathsf{P}_{t-1}(R_t < VaR_t(p)) = p$. If we define the hit sequence of $VaR_t(p)$ violations as

$$I_t = \begin{cases} 1, \text{ if } \; R_t < VaR_t(p), \\ \\ 0, \text{ else} \end{cases}$$

then Christoffersen (1998) tests the null hypothesis that

$$H_0 : I_t \sim i.i.d : Bernoull(p)$$

against

$$H_1 : I_t \sim i.i.d : Bernoull(\bar{p})$$

which is a test that on average the coverage is correct. This test can be performed using the sign procedure that we propose here. Under H_0 the likelihood function of the sequence of hit is given by

$$L_0(I_1, ...I_T, p) = \prod_{t=1}^{T} p^{I_t}(1-p)^{1-I_t} = p^{S_T}(1-p)^{n-S_T},$$

where $S_T = \sum_{t=1}^{T} I_t$, and under the alternative H_1 this function takes the form

$$L_1(I_1, ...I_T, \bar{p}) = \bar{p}^{S_T}(1-\bar{p})^{n-S_T}.$$

Thus, the test statistic for testing H_0 against H_1 is given by:

$$S_T = \sum_{t=1}^{T} I_t$$

(2004).

where under H_0 S_T follows a binomial distribution $Bi(T, p)$.

4.2.2 Point-optimal sign test for a non constant hypothesis

Now, let $y = (y_1, ..., y_n)'$ be an observable $n \times 1$ vector of independent random variables such that $\mathsf{P}[y_t \geq 0] = p_t$, for $t = 1, ...n$, and suppose we wish to test

$$H_0 : \mathsf{P}[s(y_t) = 1] = p_{t,0}, \; t = 1, ..., n, \tag{4.9}$$

against

$$H_1 : \mathsf{P}[s(y_t) = 1] = p_{t,1}, \; t = 1, ..., n. \tag{4.10}$$

Again for simplicity of exposition we assume that $p_{t,0}, p_{t,1} \neq 0, 1$.

Theorem 1 *Under assumption (4.1) the test with critical region*

$$C = \{(y_1, ..., y_n) : \sum_{t=1}^{n} \ln[\frac{p_{t,1}(1 - p_{t,0})}{p_{t,0}(1 - p_{t,1})}]s(y_t) > c_1\}$$

is the best point-optimal sign test for the hypothesis (4.9) against the alternative (4.10). The value of c_1 is chosen so that

$$\mathsf{P}((y_1, ..., y_n) \in C \mid H_0) \leq \alpha,$$

where α is an arbitrary significance level.

We use the same steps as in subsection (4.2.1) to prove Theorem 1. The test statistic is given by:

$$S_n^* = \sum_{t=1}^{n} a_t(0 \mid 1)s(y_t), \tag{4.11}$$

where

$$a_t(0 \mid 1) = \ln[\frac{p_{t,1}(1 - p_{t,0})}{p_{t,0}(1 - p_{t,1})}].$$

Contrary to the results in the pervious subsection, the test that maximized the power against a particular alternative $p_{t,1}$ depends on this alternative. Some additional principal has to be introduced to choose the optimal alternative that maximize the power of POS test. In the special case of $p_{t,0} = p_0$ and $p_{t,1} = p_1$, where p_0 and p_1 are constants, the test statistic (4.11) corresponds to *uniformly most powerful* (UMP) test based on $s(y_1), ..., s(y_n)$.

4.3 Sign-based tests in linear and nonlinear regressions

In the presence of some types of heteroskedasticity, the parametric tests proposed to improve inference may exhibit poor size control and/or low power. For example, when there is a break in the disturbances' variance, simulation results show that usual tests based on White's (1980) correction of the variance, which is supposed to be robust against heteroskedasticity, have very low power. On the other hand, many *exact* parametric tests developed in the literature typically assume normal disturbance. The latter assumption may be unrealistic and, in presence of heavy tails and asymmetric distributions, simulation studies show that these tests may not do very well in terms of power. Furthermore, the statistical procedures developed for inference on the parameters of *nonlinear* models are typically based on asymptotic approximations and there are few exact inference methods outside the linear model framework. This section proposes exact simple optimal sign-based tests to test the parameter values in linear and nonlinear regression models. These tests are valid under weak distributional assumptions such as heteroscedasticity of unknown form and non-normality. We propose a test for the null hypothesis that a vector of coefficients in a linear model is zero. We also derive a test for the null that a vector of coefficients in linear or nonlinear model is equal to an arbitrary constant vector.

4.3.1 Testing zero coefficient hypothesis in linear models

Let $y = (y_1, ..., y_n)'$ be an observable $n \times 1$ vector of independent random variables. Suppose that the variable y_t can be linearly explained by a variable x_t :

$$y_t = \beta' x_t + \varepsilon_t, \ t = 1, ..., n. \tag{4.12}$$

where $\beta \in \mathbb{R}^k$ is an unknown vector of parameters and ε_t is a disturbance variable such that

$$\varepsilon_t \mid X \sim F_t(. \mid X) \tag{4.13}$$

and

$$\mathsf{P}[\varepsilon_t \geq 0 \mid X] = \mathsf{P}[\varepsilon_t < 0 \mid X] = \frac{1}{2} \tag{4.14}$$

where $X = [x_1, ..., x_n]'$ is an $n \times k$ matrix. Suppose that we wish to test

$$H_0 : \beta = 0.$$

against

$$H_1 : \beta = \beta_1. \tag{4.15}$$

The likelihood function of the sample $\{y_t\}_{t=1}^{n}$ is given by

$$L(U(n), \beta, X) = \prod_{t=1}^{n} \mathsf{P}[y_t \geq 0 \mid X]^{s(y_t)} (1 - \mathsf{P}[y_t \geq 0 \mid X])^{1-s(y_t)}$$

where

$$\mathsf{P}[y_t \geq 0 \mid X] = 1 - \mathsf{P}[\varepsilon_t < -\beta' x_t \mid X].$$

Under H_0 we have,

$$\mathsf{P}[y_t \geq 0 \mid X] = 1 - \mathsf{P}[\varepsilon_t < 0 \mid X] = \frac{1}{2}$$

and, under the alternative H_1

$$\mathsf{P}[y_t \geq 0 \mid X] = 1 - \mathsf{P}[\varepsilon_t < -\beta_1' x_t \mid X]. \tag{4.16}$$

Based on Theorem 1 and from the value of $\mathsf{P}[y_t \geq 0 \mid X]$ under H_0 and H_1, we deduce the following result.

Proposition 2 *Under assumptions (4.1) and (4.14), the best point-optimal conditional sign test for the hypothesis H_0 against H_1 rejects H_0 when*

$$\sum_{t=1}^{n} a_t(0 \mid 1) s(y_t) > c_1(\beta_1)$$

where, for $t = 1, ..., n$,

$$a_t(0 \mid 1) = \ln\left[\frac{1}{\frac{1}{1-\mathsf{P}[\varepsilon_t \leq -\beta_1' x_t \mid X]} - 1}\right].$$

The value of $c_1(\beta_1)$ is chosen such that

$$\mathsf{P}\left(\sum_{t=1}^{n} a_t(0 \mid 1) s(y_t) > c_1(\beta_1) \mid H_0\right) \leq \alpha$$

where α is an arbitrary significance level.

Note that the point-optimal conditional sign test given by Proposition 2 controls size for any distribution of the error term which satisfies our assumption that the median equal zero. Under H_0 the test is distribution-free and allows for heteroskedasticity of unknown form. However, under H_1 the test statistic will depend on the form of the distribution function of the error term. Consequently, the power function of the POS test will depend on distribution of ε_t. In what follows, we consider that under H_1 the disturbances follow a homoskedastic Normal distribution. In other words, we substitute the optimal weights $a_t(0 \mid 1)$ by weights derived from the normal distribution. This may affect the power of POS test. However, the simulation study shows that there is almost no loss in terms of power when we misspecify the distribution function of ε_t [see Tables 7-8]. If we consider that under H_1

$$\varepsilon_t \sim \mathcal{N}(0, 1).$$

then the test statistic is given by

$$S_n^*(\beta_1) = \sum_{t=1}^{n} a_t(0 \mid 1)s(y_t) \tag{4.17}$$

where, for $t = 1, ..., n$,

$$a_t(0 \mid 1) = \ln\left[\frac{1}{\frac{1}{\Phi(\beta_1' x_t)} - 1}\right]. \tag{4.18}$$

where $\Phi(.)$ represents the CDF of normal distribution. To implement the *POS* test derived above, we compute the quantiles of the random variables (4.17). To simulate (4.17) we need to generate a sequence of $\{s(y_t)\}$ under H_0. In particular we need a sequence of $\{s(\varepsilon_t)\}$ satisfying (4.14). Since the variable $s(\varepsilon_t)$ takes only two values 0 and 1, the computation of the test statistic (4.17) reduces to generating a sequence of Bernoulli random variables of given length with subsequent summation with the corresponding weights (4.18). We now describe the algorithm to implement the point-optimal conditional sign test:

1. compute the test statistic $S_n^*(\beta_1)^0$ based on the observed data;
2. generate a sequence of Bernoulli random variables $\{s(\varepsilon_i)\}_{i=1}^n$ satisfying (4.14);
3. compute $S_n^*(\beta_1)^j$ using the generated sequence $\{s(\varepsilon_i)\}_{i=1}^n$ and the corresponding weights $\{a_i(0 \mid 1)\}_{i=1}^n$;
4. choose B such that $\alpha(B+1)$ is an integer and repeat steps $(1) - (3)$ B times;
5. compute the $(1-\alpha)\%$ quantile, denoted $c(\beta_1)$, of the sequence $\{S_n^*(\beta_1)^j\}_{j=1}^B$;
6. reject the null hypothesis at level α if $S_n^*(\beta_1)^0 \geq c(\beta_1)$.

4.3.2 Testing the general hypothesis $\beta = \beta_0$ in linear and nonlinear models

Now let us consider the following general model:

$$y_t = f(x_t, \beta) + \varepsilon_t, \; t = 1, ..., n, \tag{4.19}$$

where $f(.)$ is a scalar function, $\beta \in \mathbb{R}^k$ is an unknown vector of parameters, and ε_t is a disturbance such that (4.13) and (4.14). Suppose we wish to test

$$H_0 : \beta = \beta_0 \tag{4.20}$$

against

$$H_1 : \beta = \beta_1 \; .$$

The test of H_0 against H_1 can be constructed in the same way as in the previous subsection. We first need to transform equation (4.19) such that we can find the same structure as before. The model (4.19) is equivalent to the following transformed model

$$\tilde{y}_t = g(x_t, \beta, \beta_0) + \varepsilon_t$$

where

$$\tilde{y}_t = y_t - f(x_t, \beta_0) \; \text{ and } \; g(x_t, \beta, \beta_0) = f(x_t, \beta) - f(x_t, \beta_0).$$

For simplicity of exposition, in the rest of this section we focus on the linear case where $f(x_t, \beta) = \beta' x_t$. We deal with the nonlinear case in the appendix. We have,

$$\tilde{y}_t = \widetilde{\beta}' x_t + \varepsilon_t.$$

where

$$\tilde{y}_t = y_t - \beta_0' x_t \text{ and } g(x_t, \beta, \beta_0) = \widetilde{\beta}' x_t = (\beta - \beta_0)' x_t.$$

The hypothesis testing (4.20) is equivalent to test

$$\bar{H}_0 : \tilde{\beta} = 0,$$

against

$$\bar{H}_1 : \tilde{\beta} = \tilde{\beta}_1 = \beta_1 - \beta_0.$$

Consider the following vector of signs

$$\tilde{U}(n) = [s(\tilde{y}_1), ..., s(\tilde{y}_n)]',$$

where, for $t = 1, ..., n$,

$$s(\tilde{y}_t) = \begin{cases} 1, \; if \; \tilde{y}_t \geq 0 \\ \\ 0, \; if \; \tilde{y}_t < 0 \end{cases}.$$

The test of H_0 against H_1 can be derived using Theorem 1 and following the same steps as in subsection 4.3.1. We have the following result.

Proposition 3 *Under assumptions (4.1) and (4.14), the best point-optimal conditional sign test for H_0 against H_1 rejects H_0 when*

$$\sum_{t=1}^{n} \tilde{a}_t(0 \mid 1) s(y_t - \beta_0' x_t) > c_1(\beta_1),$$

where, for $t = 1, ..., n$

$$\tilde{a}_t(0 \mid 1) = \ln\left[\frac{1}{\frac{1}{1 - \mathsf{P}[\varepsilon_t \leq -(\beta_1 - \beta_0)' x_t \mid X]} - 1}\right].$$

The value of $c_1(\beta_1)$ *is chosen so that*

$$\mathsf{P}(\sum_{t=1}^{n} \tilde{a}_t(0 \mid 1)s(y_t - \beta_0' x_t) > c_1(\beta_1) \mid H_0) \leq \alpha,$$

and α *is an arbitrary significance level.*

If under H_1 $\varepsilon_t \sim \mathcal{N}(0, 1)$, then the test statistic is given by:

$$S_n^*(\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(0 \mid 1)s(y_t - \beta_0' x_t) \tag{4.21}$$

where

$$\tilde{a}_t(0 \mid 1) = \ln\left[\frac{1}{\frac{1}{\Phi((\beta_1-\beta_0)' x_t)} - 1}\right], \; t = 1, \ldots, n. \tag{4.22}$$

4.4 Power envelope and the choice of the optimal alternative

We study the power properties of the POS test. We derive the power envelope and analyze the impact of the choice of the alternative hypothesis β_1 on the power function. Since the POS test depends on the alternative hypothesis, we propose an approach, called the adaptive approach, to choose an alternative β_1 such that the power curve of the POS test is close to the power envelope curve.

4.4.1 Power envelope of the point-optimal sign test

We derive the upper bound of the power function of the POS test (hereafter the power envelope). One advantage of point-optimal tests is they can be used to trace out the maximum attainable power for a given testing problem. This power envelope provides a natural benchmark against which test procedures can be compared. The POS test optimizes power at given point of the parameter space. The test statistic is a function of

β_1,

$$S_n^*(\beta_1) = \sum_{t=1}^{n} a_t(0 \mid 1)s(y_t)$$

where

$$a_t(0 \mid 1) = \ln\left[\frac{1}{\frac{1}{1-\mathsf{P}[\varepsilon_t \le -\beta_1' x_t | X]} - 1}\right].$$

Its power function is also a function of β_1 and it is given by:

$$\Pi(\beta, \beta_1) = \mathsf{P}[S_n^*(\beta_1) > c_1]$$

where c_1 satisfies

$$\mathsf{P}[S_n^*(\beta_1) > c_1 \mid H_0] \le \alpha.$$

Theorem 4 *Under assumptions (4.1) and (4.14), the power function of the POS test at given point β_1 is given by*

$$\Pi(\beta, \beta_1) = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \frac{I(u)}{u} du,$$

where, for $u \in \mathbb{R}$,

$$I(u) = (\frac{1}{2})^n \operatorname{Im}\left\{\prod_{t=1}^{n}[\exp(-iu\frac{c_1}{n}) + \exp(iu(\, a_t(0 \mid 1) - \frac{c_1}{n}\,))]\right\}$$

and, for $t = 1, \ldots, n$,

$$a_t(0 \mid 1) = \ln\left[\frac{1}{\frac{1}{1-\mathsf{P}[\varepsilon_t \le -\beta_1' x_t | X]} - 1}\right].$$

$i = \sqrt{-1}$, $\operatorname{Im}\{z\}$ denotes the imaginary part of a complex number z, and the value of c_1 is chosen so that

$$\mathsf{P}[S_n^*(\beta_1) > c_1 \mid H_0] \le \alpha$$

where α is an arbitrary significance level.

Since the test statistic $S_n^*(\beta_1)$ is optimal against an alternative β_1, the envelope power function, denoted $\bar{\Pi}(\beta)$, is a function that associates the value $\Pi(\beta, \beta_1)$ to each element $\beta \in \mathbb{R}^k$,

$$\bar{\Pi}(\beta) = \Pi(\beta, \beta) = \mathsf{P}[S_n^*(\beta) > c_1]. \tag{4.23}$$

The objective is to find a value of β_1 at which the power curve of the *POS* test remains close to the relevant power envelope. For a given value Π of the power function and level α of the *POS* test, one can find an alternative $\beta_1(\Pi, \alpha)$ by inverting the power envelope function $\bar{\Pi}(\beta)$. Thus, for any given value $\Pi \in [\alpha, 1]$, the family of *POS* test statistics can be written as follows

$$S_n^*(\Pi) = \sum_{t=1}^{n} a_t(0 \mid 1) s(y_t),$$

where

$$a_t(0 \mid 1) = \ln\left[\frac{1}{\frac{1}{1-\mathsf{P}[\varepsilon_t \leq -\beta_1(\Pi,\alpha)' x_t \mid X]} - 1}\right].$$

Although every member of this family is admissible, it is possible that some values of Π may yield tests whose power functions lie close to the power envelope over a considerable range. Past research suggests that values of Π near one-half often have this property, see for example King (1988), Dufour and King (1991), and Elliot, Rothenberg and Stock (1996). Consequently, one can choose as an optimal alternative the one which corresponds to $\Pi = 0.5$.

Based on Theorem 4 and equation (4.23), the value of β_1 corresponding to $\Pi = 0.5$ is the solution of the following equation[4]

$$\int_0^{\infty} \operatorname{Im}\left\{ \frac{\prod_{t=1}^{n}\left[\exp(-iu\frac{c_1}{n}) + \exp(iu(a_t(0 \mid 1) - \frac{c_1}{n}))\right]}{u} \right\} du = 0. \tag{4.24}$$

[4]Using the properties of the cumulative density function (monotonically increasing, continuous $\lim_{c \to -\infty} \Pr(z < c) = 0$, and $\lim_{c \to +\infty} P_t(z < c) = 1$) one can show that equation (4.24) has a unique solution.

In practice, an exact solution of equation (4.24) is not feasible, since it is hard to compute the expression of $\text{Im}\{I(u)\}$ and the integral $\int_0^\infty \frac{I(u)}{u} du$ is difficult to evaluate. The latter can be approximated using results by Imhof (1961), Bohmann (1972), and Davies (1973), who propose a numerical approximation of the distribution function using the characteristic function. The proposed approximation introduces two types of errors: discretization and truncation errors. Davies (1973), proposes a criterion to control for discretization error and Davies (1980) proposes three different bounds to control for truncation error. Another way to solve the power envelope function for β_1 is to use simulations. One can use simulations to approximate the power envelope function and calculate the optimal alternative which corresponds to the value of $\bar{\Pi}(\beta_1)$ near one-half.

Let us now examine the impact of the choice of the alternative β_1 on the power function. In what follows, we use simulations to plot the power curves of the POS test under different alternatives and compare them to the power envelope. We find the following results:

Insert Figures 2-7.

The above Figures compare the power curves of the POS test under different alternatives to the power envelope for different data generating processes (hereafter DGP's). We consider a linear regression model with one regressor and the error terms follow one of the following distributions: Normal, Cauchy, mixture of Normal and Cauchy, Normal with GARCH(1,1) and jump, Normal with non-stationary $GARCH(1,1)$, and Normal with a break in variance. We describe these DGP's in more detail in section 4.6. Based on simulation results, we find that the value of the alternative β_1 affects the power function. Particularly, when the alternative is far from the null $\beta = 0$ the power curve of the POS test moves away from the power envelope curve.

Since the previous approach to finding the optimal alternative is somewhat arbitrary way, we propose a natural approach, called the adaptive approach, based on split-sample technique to estimate the optimal alternative.

4.4.2 An adaptive approach to choose the optimal alternative

Existing adaptive statistical methods use the data to determine which statistical procedure is most appropriate for a specific statistical problem. These methods are usually performed in two steps. In the first step a selection statistic is computed that estimates the shape of the error distribution. In the second step the selection statistic is used to determine an effective statistical procedure for the error distribution. More details about the adaptive statistical methods can be found in Gorman (2004).

The adaptive approach that we consider here is somewhat different from the existing adaptive statistical approaches. We propose the split-sample technique to choose an alternative β_1 such that the power curve of the *POS* test is close to the power envelope.[5] The alternative hypothesis β_1 is unknown and a practical problem consists in finding its independent estimate. To make size control easier, we estimate β_1 from a sample which is independent from the one that we use to run the *POS* test. This can be easily done by splitting the sample. The idea is to divide the sample into two independent parts and to use the first one to estimate the value of the alternative and the second one to compute the *POS* test statistic. Consider again the model given by (4.12) and let $n = n_1 + n_2$, $y = (y'_{(1)}, y'_{(2)})'$, $X = (X'_{(1)}, X'_{(2)})'$, and $\varepsilon = (\varepsilon'_{(1)}, \varepsilon'_{(2)})'$ where the matrices $y_{(i)}$, $X_{(i)}$, and $\varepsilon_{(i)}$ have n_i rows $(i = 1, 2)$. We use the first n_1 observations on y and X, respectively $y_{(1)}$ and $X_{(1)}$, to estimate the alternative hypothesis β_1 using, for example, OLS estimation method:

$$\hat{\beta}_1 = (X'_{(1)}X_{(1)})^{-1}X'_{(1)}y_{(1)}.$$

However, the OLS estimator is known to be very sensitive to outliers and non-normal errors, consequently it is important to choose a more appropriate method to estimate β_1. In presence of outliers many estimators are proposed to estimate the coefficients in regression model such that the least median of squares (LMS) estimator [Rousseeuw and Leroy (1987)], the least trimmed sum of squares (LTS) estimator [Rousseeuw (1983)],

[5] For more details about split-sample technique, the reader can consult Dufour and Torrès (1998) and Dufour and Jasiak (2001).

the S-estimators [Rousseeuw and Yohai (1984)], and the τ-estimators [Yohai and Zamar (1988)].

Because $\hat{\beta}_1$ is independent of $X_{(2)}$, one can use the last n_2 observations on y and X, respectively $y_{(2)}$ and $X_{(2)}$, to calculate the test statistic and to get a valid POS test:

$$S_n^*(\hat{\beta}_1) = \sum_{t=n_1+1}^{n} a_t(0 \mid 1) s(y_t),$$

where, for $t = n_1 + 1, \ldots, n$,

$$a_t(0 \mid 1) = \ln\left[\frac{1}{\frac{1}{1-\mathrm{P}[\varepsilon_t \leq -((X_{(1)}'X_{(1)})^{-1}X_{(1)}'y_{(1)})'x_t \mid X]} - 1}\right].$$

Note that different choices for n_1 and n_2 are clearly possible. Alternatively, one could select randomly the observations assigned to the vectors $y_{(1)}$ and $y_{(2)}$. As we will show latter 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 one can get a more powerful test once we use a relatively small number of observations for computing the alternative hypothesis and keep more observations for the calculation of the test statistics. This point is illustrated below by simulation experiments. We use simulations to compare the power curves of the split-sample-based POS test (hereafter SS-POS test) to the power envelope (hereafter PE) under different split-sample sizes and for different DGP's. We use the same DGP's as those we have considered in the last subsection We find the following results:

Insert Figures 8-13.

From the above figures, we see that using approximately 10% of sample to estimate the alternative yields a power which is typically very close to the power envelope. This is true for all DGP's that we have considered in our simulation study.

4.5 Point-optimal sign-based confidence regions

We will briefly describe how we can build confidence regions, say $C_{\beta}(\alpha)$, for a vector of unknown parameters β, with known level α, using *POS* tests. Consider the following model:

$$y_t = \beta' x_t + \varepsilon_t, \; t = 1, ..., n,$$

where $\beta \in \mathbb{R}^k$ is an unknown vector of parameters and ε_t is a disturbance such that (4.13) and (4.14). Suppose we wish to test

$$H_0 : \beta = \beta_0$$

against

$$H_1 : \beta = \beta_1. \tag{4.25}$$

The idea consists to find all the values of $\beta_0 \in \mathbb{R}^k$ such that,

$$S_n^{*(0)}(\beta_1) = \sum_{t=1}^{n} \left\{ \ln \left[\frac{1}{\frac{1}{1-\mathsf{P}[\varepsilon_t \leq -(\beta_1-\beta_0)' x_t | X]} - 1} \right] s(y_t - \beta_0' x_t) \right\} < c(\beta_1),$$

where $S_n^{*(0)}(\beta_1)$ is the observed value of $S_n^*(\beta_1)$. The critical values for the former test are found by solving

$$\mathsf{P}[S_n^*(\beta_1) > c(\beta_1) \mid \beta = \beta_0] \leq \alpha.$$

Thus, the confidence region $C_{\beta}(\alpha)$ can be defined as follows:

$$C_{\beta}(\alpha) = \left\{ \beta_0 : S_n^{*(0)}(\beta_1) < c(\beta_1), \text{ for } \mathsf{P}[S_n^*(\beta_1) > c(\beta_1) \mid \beta = \beta_0] \leq \alpha \right\}.$$

Moreover, given the confidence region $C_{\beta}(\alpha)$, one can derive confidence intervals for the components of the vector β using the projection techniques.[6] The latter can be used to

[6] More details about the projection technique can be find in Dufour (1997), Abdelkhalek and Dufour (1998), Dufour and Kiviet (1998), Dufour and Jasiak (2001), and Dufour and Taamouti (2005).

find confidence sets, say $g(C_\beta(\alpha))$. for general transformations g of β in $\mathbb{R}^m$. Since

$$\beta \in C_\beta(\alpha) \Rightarrow g(\beta) \in g(C_\beta(\alpha)), \tag{4.26}$$

for any set $C_\beta(\alpha)$, we have:

$$\mathsf{P}[\beta \in C_\beta(\alpha)] \geq 1 - \alpha \Rightarrow \mathsf{P}[g(\beta) \in g(C_\beta(\alpha))] \geq 1 - \alpha, \tag{4.27}$$

where

$$g(C_\beta(\alpha)) = \{\delta \in \mathbb{R}^m : \exists \beta \in C_\beta(\alpha),\ g(\beta) = \delta\}.$$

Given (4.26)-(4.27), $g(C_\beta(\alpha))$ is a conservative confidence region for $g(\beta)$ with level $1-\alpha$. If $g(\beta)$ is a scalar, then we have

$$\mathsf{P}\left[inf\,\{g(\beta_0), \text{for } \beta_0 \in C_\beta(\alpha)\} \leq g(\beta) \leq sup\,\{g(\beta_0),\ \text{for } \beta_0 \in C_\beta(\alpha)\}\right] > 1 - \alpha.$$

4.6 Monte Carlo study

We present simulation results illustrating the performance of the procedures given in the preceding sections. Since the number of tests and alternative models is large, we have limited our results to two groups of data generating processes (DGP) which correspond to different forms of symmetric and asymmetric distributions and different forms of heteroskedasticity.

4.6.1 Size and Power

To assess the performance of the POS test, we run a simulation study to compare its size and power to those of some common tests under various general DGP's. We choose our DGP's to illustrate performance in different contexts that one can encounter in practice.

The model under consideration is given by:

$$y_t = \beta x_t + \varepsilon_t, \ t = 1, ..., n, \tag{4.28}$$

where β is an unknown parameter and the disturbances ε_t are independent and can follow different distributions. We wish to test

$$H_0 : \beta = 0.$$

Let us now specify the DGP's that we consider in the simulation study. The first group of DGP's that we examine represents different forms of symmetric and asymmetric distributions of the error terms:

1. Normal:

$$\varepsilon_t \sim \mathcal{N}(0, 1), t = 1, ..., n.$$

2. Cauchy:

$$\varepsilon_t \sim Cauchy, t = 1, ..., n.$$

3. Student:

$$\varepsilon_t \sim Student(2), t = 1, ..., n.$$

4. Mixture:

$$\varepsilon_t \sim s_t \mid \varepsilon_t^C \mid -(1 - s_t) \mid \varepsilon_t^N \mid, t = 1, ..., n,$$

with

$$\mathsf{P}[s_t = 1] = \mathsf{P}[s_t = 0] = \frac{1}{2} \text{ and } \varepsilon_t^C \sim Cauchy, \ \varepsilon_t^N \sim \mathcal{N}(0, 1).$$

The second group of DGP's that we consider represents different forms of heteroskedasticity:

5. Break in variance:

$$\varepsilon_t \sim \begin{cases} \mathcal{N}(0,1) & \text{for } t \neq 25 \\ \\ \sqrt{1000}\mathcal{N}(0,1) & \text{for } t = 25 \end{cases}$$

6. $GARCH(1,1)$ with jump:

$$\varepsilon_t \sim \begin{cases} \mathcal{N}(0,\sigma_\varepsilon^2(t)) & \text{for } t \neq 25 \\ \\ 50\ \mathcal{N}(0,\sigma_\varepsilon^2(t)) & \text{for } t = 25 \end{cases}$$

and

$$\sigma_\varepsilon^2(t) = 0.00037 + 0.0888\varepsilon_{t-1}^2 + 0.9024\sigma_\varepsilon^2(t-1).$$

7. Non stationary $GARCH(1,1)$:

$$\varepsilon_t \sim \mathcal{N}(0,\sigma_\varepsilon^2(t)),\ t = 1,\dots,n.$$

and

$$\sigma_\varepsilon^2(t) = 0.75\varepsilon_{t-1}^2 + 0.75\sigma_\varepsilon^2(t-1)$$

In this case we run two different simulations which correspond to two different initial values of $\sigma_\varepsilon^2(t)$. Figure 19 corresponds to $\sigma_\varepsilon^2(0) = 0.2$ and figure 20 to $\sigma_\varepsilon^2(0) = 0.0002$.

8. Exponential variance:

$$\varepsilon_t \sim \mathcal{N}(0,\sigma_\varepsilon^2(t)),\ t = 1,\dots,n$$

and

$$\sigma_\varepsilon(t) = \exp(0.5t),\ t = 1,\dots,n.$$

The explanatory variable x_t is generated from a mixture of normal and χ^2 distributions, and all simulated samples are of size $n = 50$. We perform $M1 = 10000$ simulations to evaluate the probability distribution of the POS test statistic and $M2 = 5000$ simulations to estimate the power functions of the POS test and these of other tests.

4.6.2 Results

We compare the power envelope curve to the power curves of the 10% split-sample *POS* test, the *t-test* (or *CT-test*)[7], the sign test proposed by Campbell and Dufour (1995) (hereafter *CD(1995)* test)[8], and the *t-test* based on White's (1980) correction of variance (hereafter *WT-test or CWT-test*)[9]. The simulation results are given in Tables 1-6 and Figures 14-22. These results correspond to different DGP's that we have described before. Tables 1-6 compare the power envelope, the *POS* test, the *t-test, the CD(1995)* test, and the *WT-test* under different split-sample sizes and alternative hypotheses. Figures 14-22 compare the power envelope curve to these of the 10% split-sample *POS* test, *CD(1995)* test, the *t-test (or CT-test)*, and the *WT-test (or CWT-test)*.

Table 1 and Figure 14 correspond to the case where the error terms follow normal distribution. Table 1 shows the power function depends on the alternative hypothesis. When β_1 is far from the null, the power curve moves away from the power envelope curve [see also Figure 2]. When we use the split-sample technique to choose the alternative hypothesis, we see that using approximately 10% of the sample to estimate β_1 yields a power which is typically very close to the power envelope. Figure 14 shows that the *t-test* is more powerful than the *CWT-test, the* 10% split-sample *POS* test, and *CD(1995)* test. This is an expected result, since under normality the *t-test* is the most powerful test. However, the power curve of 10% split-sample *POS* test is still very close to the power envelope and do better than the *CD(1995)* test. We also note that the *t-test* based on White's (1980) correction of variance does not control size. The last column of Table 1 gives the power of *WT-test* after size correction.

Table 2 and Figure 15 correspond to the Cauchy distribution. From table 2, we see that the power of the *POS* test depends again on the alternative hypothesis that we

---

[7] *CT-test* corresponds to the power of *t-test* after size correction. Under some DGP's the *t-test* may not control its size, thus we adjust the power function such that *CT-test* controls its size.

[8] The sign test of Campbell and Dufour (1995) has discrete distribution and it is not possible (without randomization) to obtain test whose size is precisely 5%; here the size of this test is 5.95% for $n = 50$.

[9] *CWT-test* corresponds to the power of *WT-test* after size correction. Under some DGP's the *WT-test* may not control its size, thus we adjust the power function such that *CWT-test* controls its size.

consider. In particular, when the value of β_1 is far from the null, the power curve moves away from the power envelope curve. We also note that using approximately 10% of sample to estimate β_1 yields a power which is typically very close to the power envelope. Figure 15 shows that the 10% split-sample *POS* test is more powerful than the *CD(1995)* test, the *t-test*, the *WT-test*, and it still close to the power envelope.

Tables 3, 5, and 6 and Figures 16, 18, and 19-20 correspond to the Mixture, GARCH(1,1) with jump, and Non stationary GARCH(1,1) cases, respectively. We get similar results, as in the Normal and Cauchy distributions, in terms of the impact of β_1 on the power function and the values n_1 and n_2 that we have to consider. Figures 16, 18, and 19-20 show that the 10% split-sample *POS* test is more powerful than the *WT-test, the CD(1995)* test, the *t-test*, and is very close to the power envelope. For the mixture error terms, the *WT-test* and *the t-test* do not control size, thus we adjust the power function such that these tests control their size. Table 4 and Figure 17 correspond to the Break in variance case. As we can see the power curve of the *t-test* and the *WT-test* are almost flat, whereas the 10% split-sample *POS* test does very well and is more powerful than the *CD(1995) test.* Finally, for the Student case, Figure 21 shows that 10% split-sample *POS* test is more powerful than the *CD (1995)* test and the *t-test.*

From the above results, we draw the following conclusions. First, it is clear that the choice of the alternative β_1 has an impact on the power function of the *POS* test. Second, the adaptive approach based on split-sample technique allows one to choose an optimal value of the alternative β_1. We should use a small part, approximately 10%, of the sample to estimate the alternative and the rest to calculate the test statistic. Third, for *DGP's* with normal and heteroskedastic disturbances, the power curve of 10% split-sample *POS* test is close to the power envelope. However, for non-normal disturbances the power curve of the 10% split-sample *POS* test is somewhat far from the power envelope. Finally, except for the Normal distribution, all simulations results show that the 10% split-sample *POS* test performs better than the *CD(1995)* test, the *t-test*, and the *WT-test* (including *CT-test* and the *WCT-test).*

We also run simulations to compare the power of the 10% split-sample POS test calculated under the true weights $a_t(0 \mid 1)$ with that of the 10% split-sample POS test calculated using normal weights. The results are given in tables 7 and 8. We see that by using the true weights one may improve the power of the 10% split-sample POS test. However, the power loss when we substitute the true weights by normal weights is still very small.

4.7 Conclusion

In this chapter, we have proposed an exact and simple conditional sign-based point-optimal test to test the parameters in linear and nonlinear regression models. The test is distribution-free, robust against heteroskedasticity of an unknown form, and it may be inverted to obtain confidence sets for the vector of unknown parameters. Since the point-optimal conditional sign test maximizes the power at a given value of the alternative, we propose an approach based on split-sample technique to choose an alternative such that the power curve of the point-optimal conditional sign test is close to that of power envelope. Our simulation study shows that by using approximately 10% of sample to estimate the alternative hypothesis and the rest to calculate the test statistic, the power curve of the proposed "quasi" point-optimal conditional sign test is typically close to the power envelope curve.

To assess the performance of the point-optimal conditional sign tests, we run a simulation study to compare its size and power to those of some usual tests under various general DGP's. We consider different DGP's to illustrate different contexts that one can encounter in practice. These DGP's are relative to the non-normal, asymmetric, and heteroskedastic disturbances. The results show that the 10% split-sample point-optimal conditional sign test performs better than the *t-test, the* Campbell and Dufour's (1995) sign test, and the t-test with White's (1980) variance correction.

4.8 Appendix: Proofs

Proof of Theorem 1. For our statistical problem, the likelihood function of the sample $\{y_t\}_{t=1}^n$ is given by:

$$L(U(n), p_t) = \prod_{t=1}^{n} \mathsf{P}[y_t \geq 0]^{s(y_t)}(1 - \mathsf{P}[y_t \geq 0])^{1-s(y_t)}.$$

Under H_0 this function has the form

$$L_0(U(n), p_{t,0}) = \prod_{t=1}^{n} \{p_{t,0}^{s(y_t)}(1 - p_{t,0})^{1-s(y_t)}\}$$

and under the alternative H_1 it takes the form

$$L_1(U(n), p_{t,1}) = \prod_{t=1}^{n} \{p_{t,1}^{s(y_t)}(1 - p_{t,1})^{1-s(y_t)}\}.$$

The likelihood ratio is given by:

$$\frac{L_1(U(n), p_{t,1})}{L_0(U(n), p_{t,0})} = \prod_{t=1}^{n} \{(\frac{p_{t,1}}{p_{t,0}})^{s(y_t)}\} \prod_{t=1}^{n} \{(\frac{1 - p_{t,1}}{1 - p_{t,0}})^{1-s(y_t)}\}. \tag{4.29}$$

For simplicity of exposition we suppose that $p_{t,0}, p_{t,1} \neq 0, 1$. From (4.29), the log-likelihood ratio is given by:

$$\begin{aligned} \ln\{\frac{L_1(U(n), p_1)}{L_0(U(n), p_0)}\} &= \sum_{t=1}^{n} \{s(y_t) \ln(\frac{p_{t,1}}{p_{t,0}}) + [1 - s(y_t)] \ln(\frac{1 - p_{t,1}}{1 - p_{t,0}})\} \\ &= \sum_{t=1}^{n} [q_t(1) - q_t(0)]s(y_t) + \sum_{t=1}^{n} q_t(0). \end{aligned}$$

where

$$q_t(1) = \ln(\frac{p_{t,1}}{p_{t,0}}), \quad q_t(0) = \ln(\frac{1 - p_{t,1}}{1 - p_{t,0}}).$$

The log-likelihood ratio can also be written as follows:

$$\ln\{\frac{L_1(U(n), p_1)}{L_0(U(n), p_0)}\} = \sum_{t=1}^{n} a_t(0 \mid 1)s(y_t) + b(n),$$

where

$$a_t(0 \mid 1) = q_t(1) - q_t(0), \ b(n) = \sum_{t=1}^{n} q_t(0).$$

Thus, based on the Neyman-Pearson lemma [see e.g. Lehmann (1959, p.65)], the best test of H_0 against H_1 rejects H_0 when

$$\sum_{t=1}^{n} \ln[\frac{p_{t,1}(1-p_{t,0})}{p_{t,0}(1-p_{t,1})}]s(y_t) + b(n) > c.$$

or equivalently when

$$\sum_{t=1}^{n} \ln[\frac{p_{t,1}(1-p_{t,0})}{p_{t,0}(1-p_{t,1})}]s(y_t) > c_1 \equiv c - b(n).$$

■

Proof: *POS* test in the context of nonlinear regression function. Consider the following nonlinear model,

$$y_t = f(x_t, \beta) + \varepsilon_t. \tag{4.30}$$

Suppose that we wish to test

$$H_0 : \beta = \beta_0, \tag{4.31}$$

against

$$H_1 : \beta = \beta_1. \tag{4.32}$$

The model (4.30) is equivalent to the following transformed model,

$$\tilde{y}_t = g(x_t, \beta, \beta_0) + \varepsilon_t,$$

where

$$\tilde{y}_t = y_t - f(x_t, \beta_0), \quad g(x_t, \beta, \beta_0) = f(x_t, \beta) - f(x_t, \beta_0).$$

Note that, under assumption (4.1) and conditional on X, we have

$$\tilde{y}_t, \ t = 1, ..., n, \text{ are independent.}$$

The hypothesis testing (4.31)-(4.32) is equivalent to testing

$$\bar{H}_0 : g(x_t, \beta, \beta_0) = 0, \ t = 1, ...n,$$

against

$$\bar{H}_1 : g(x_t, \beta, \beta_0) = g(x_t, \beta_1, \beta_0) = f(x_t, \beta_1) - f(x_t, \beta_0), \ t = 1, ...n,$$

The likelihood function of our sample is given by,

$$L(\tilde{U}(n), \beta, X) = \prod_{t=1}^{n} \mathsf{P}[\tilde{y}_t \geq 0 \mid X]^{s(\tilde{y}_t)}(1 - \mathsf{P}[\tilde{y}_t \geq 0 \mid X])^{1-s(\tilde{y}_t)},$$

where

$$\tilde{U}(n) = (s(\tilde{y}_1), ..., s(\tilde{y}_n))'.$$

and

$$s(\tilde{y}_t) = \begin{cases} 1, \ if \ \tilde{y}_t \geq 0 \\ 0, \ if \ \tilde{y}_t < 0 \end{cases}, \ t = 1, ..., n.$$

Under H_0 we have,

$$L_0(\tilde{U}(n), \beta_0, X) = (\frac{1}{2})^n.$$

and under H_1,

$$L_1(\tilde{U}(n), \beta_1, X) = \prod_{t=1}^{n} \mathsf{P}[\varepsilon_t \geq -g(x_t, \beta_1, \beta_0) \mid X]^{s(\tilde{y}_t)}(1 - \mathsf{P}[\varepsilon_t \geq -g(x_t, \beta_1, \beta_0) \mid X])^{1-s(\tilde{y}_t)}.$$

The log-likelihood ratio is given by

$$\ln\{\frac{L_1(\tilde{U}(n),\beta_1,X)}{L_0(\tilde{U}(n),\beta_0,X)}\} = \sum_{i=1}^{n} \tilde{a}_t(0 \mid 1)s(y_t - f(x_t,\beta_0)) + b(n),$$

where

$$\tilde{a}_t(0 \mid 1) = \ln[\frac{1}{\frac{1}{1-\mathsf{P}[\varepsilon_t \leq f(x_t,\beta_0)-f(x_t,\beta_1)|X]} - 1}],$$

and

$$\tilde{b}(n) = \sum_{i=1}^{n} \ln[\mathsf{P}[\varepsilon_t \leq f(x_t,\beta_0) - f(x_t,\beta_1) \mid X]] - \frac{n}{2}.$$

Thus, the best test of H_0 against H_1 reject H_0 when

$$\sum_{i=1}^{n} \tilde{a}_t(0 \mid 1)s(y_t - f(x_t,\beta_0)) > c_1(\beta_1),$$

where

$$\tilde{a}_t(0 \mid 1) = \ln[\frac{1}{\frac{1}{1-\mathsf{P}[\varepsilon_t \leq f(x_t,\beta_0)-f(x_t,\beta_1)|X]} - 1}],$$

and $c_1(\beta_1)$ is chosen such that

$$\mathsf{P}(\sum_{t=1}^{n} \tilde{a}_t(0 \mid 1)s(y_t - f(x_t,\beta_0) > c_1(\beta_1) \mid H_0) \leq \alpha,$$

where α is an arbitrary significance level. ■

Proof of Theorem 4. $\forall u \in \mathbb{R}$ and conditionally on X the characteristic function of $S_n^*(\beta_1)$

$$\phi_{S_n^*}(u) = E_X[\exp(iu\, S_n^*(\beta_1))] = E_X[\prod_{t=1}^{n} \exp(iu\, a_t s_t)],$$

where $a_t = a_t(0 \mid 1)$, $s(y_t) = s_t$, and $i = \sqrt{-1}$. Since y_t, for $t = 1, \ldots, n$, are independent,

$$\begin{aligned}
\phi_{S_n}(u) &= \prod_{t=1}^{n} E_X[\exp(iu\ a_t s_t)] \\
&= \prod_{t=1}^{n} \sum_{j=0}^{1} \mathsf{P}[s_t = j \mid X] \exp(iu\ a_t\ j)] \\
&= (\frac{1}{2})^n \prod_{t=1}^{n} [1 + \exp(iu\ a_t\ )].
\end{aligned}$$

According to Gil-Pelaez (1951), the conditional distribution function of $S_n^*(\beta_1)$ evaluated at c_1, for $c_1 \in \mathbb{R}$, is given by:

$$\mathsf{P}(S_n^*(\beta_1) \le c_1 \mid X) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{I(u)}{u} du, \tag{4.33}$$

where

$$I(u) = (\frac{1}{2})^n \operatorname{Im} \left\{ \prod_{t=1}^{n} [\exp(-iu\frac{c_1}{n}) + \exp(iu(\ a_t - \frac{c_1}{n}\ )] \right\}.$$

$\operatorname{Im}\{z\}$ denotes the imaginary part of a complex number z. Thus, the power function of the POS test is given by the following probability function:

$$\Pi(\beta, \beta_1) = \mathsf{P}[S_n^*(\beta_1) > c_1(\beta_1)] = 1 - \mathsf{P}[S_n^*(\beta_1) \le c_1(\beta_1)] = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \frac{I(u)}{u} du,$$

where

$$I(u) = (\frac{1}{2})^n \operatorname{Im} \left\{ \prod_{t=1}^{n} [\exp(-iu\frac{c_1}{n}) + \exp(iu(\ a_t - \frac{c_1}{n}\ )] \right\}.$$

■

Table 7: True weights versus Normal weights (Cauchy case).

		$SS-POS\ test$ [11] $with\ true\ weights$		$SS-POS\ test\ \ with\ Normal\ weights$	
β	PE	10%	20%	10%	20%
0	5.1	5.16	5.16	5.3	5.48
0.005	34.22	33.58	31.18	33.3	30.86
0.01	66.38	61.94	62.47	61.74	62.28
0.015	84.44	80.32	80.32	76.24	77.02
0.02	92.2	89.76	89.76	84.9	85.14
0.025	96.44	95.22	95.22	89.88	88.82
0.03	98.12	96.98	96.98	92.92	92.58
0.035	99	98.26	98.26	93.7	93.1
0.04	99.36	99.14	99.14	94.7	94.3
0.045	99.68	99.3	99.3	94.92	95.74
0.05	99.8	99.44	99.44	95.92	95.92
0.055	99.98	99.7	99.7	96.42	96.48
0.06	99.94	99.82	99.82	97.02	96.18
0.065	99.94	99.9	99.9	96.86	96.9

Table 8: True weights versus Normal weights (Mixture case).

		$SS-POS\ test\ \ with\ true\ weights$		$SS-POS\ test\ \ with\ Normal\ weights$	
β	PE	10%	20%	10%	20%
0	4.96	4.74	5.26	4.7	5.02
0.001	9.96	8.96	9.08	9.98	9.16
0.002	15.7	14.34	16.7	15.9	14.6
0.003	25.26	24.84	24.67	24.76	24.6
0.004	35.46	34.52	34.46	34.08	34.28
0.005	46.08	44.26	44.06	44.14	42.96
0.006	56.68	53.24	54.96	51.78	52.06
0.007	67.64	62.92	62.88	61.9	61.84
0.008	75	71.66	70.14	69.48	69.5
0.009	82.06	79.24	79.54	76.52	75.32
0.01	88.48	85.52	84.34	80.84	79.9
0.011	90.68	88.8	89.22	84.16	84.94
0.012	94.38	92.06	91.5	87.66	87.42
0.013	95.7	94.32	94.62	90.54	89.22

[11] *SS-POST*=Split-Sample POST.

Figure 1: Daily return of S&P 500 stock price index (%)

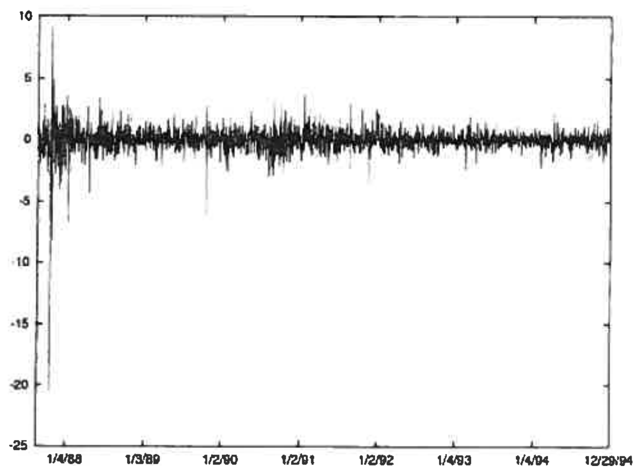


Figure 2: Power Comparison (Normal case)

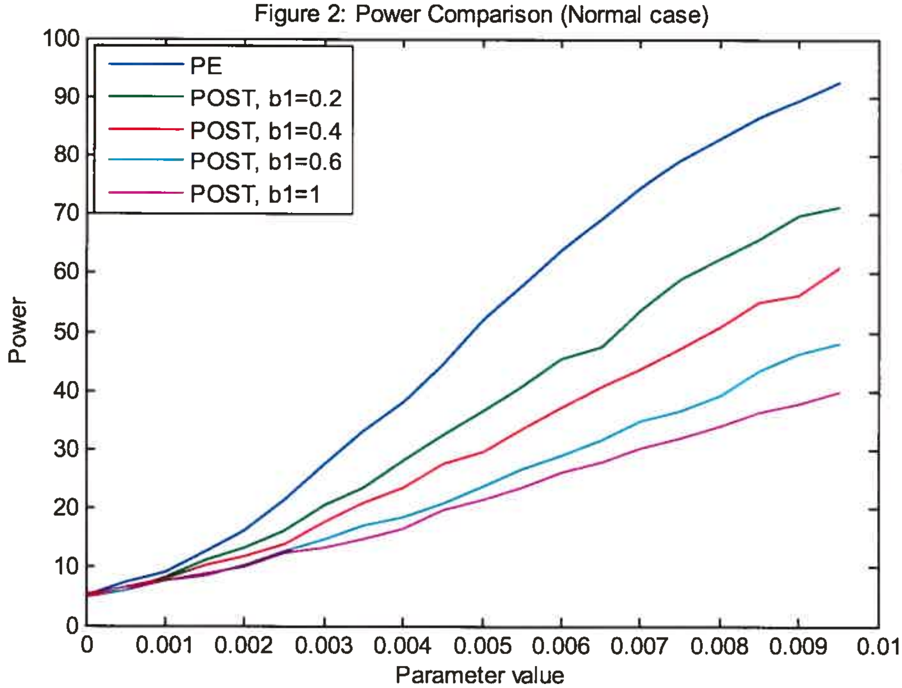


Figure 3: Power Comparison (Cauchy case)

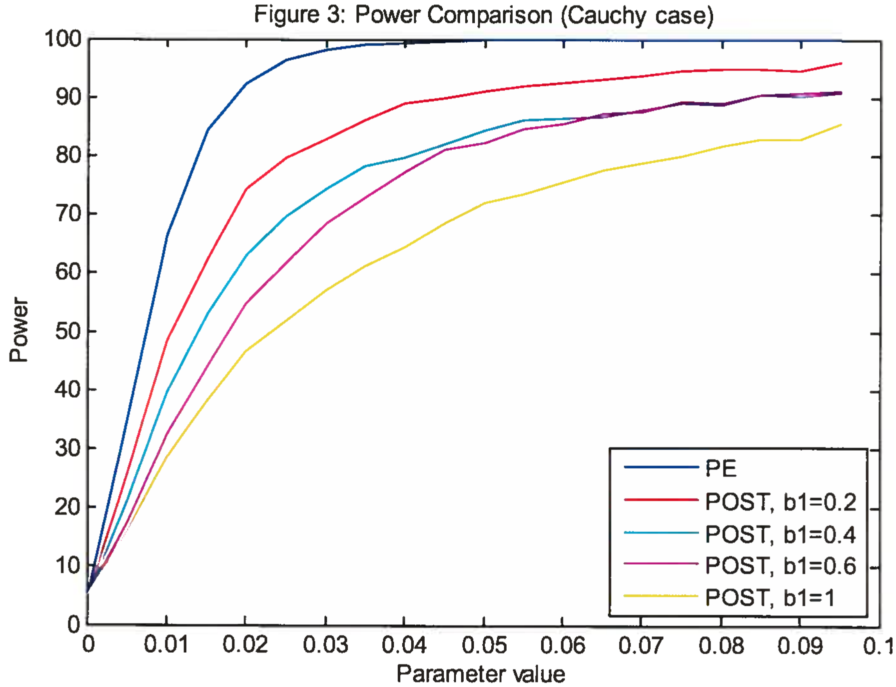


Figure 4: Power Comparison (Mixture case)

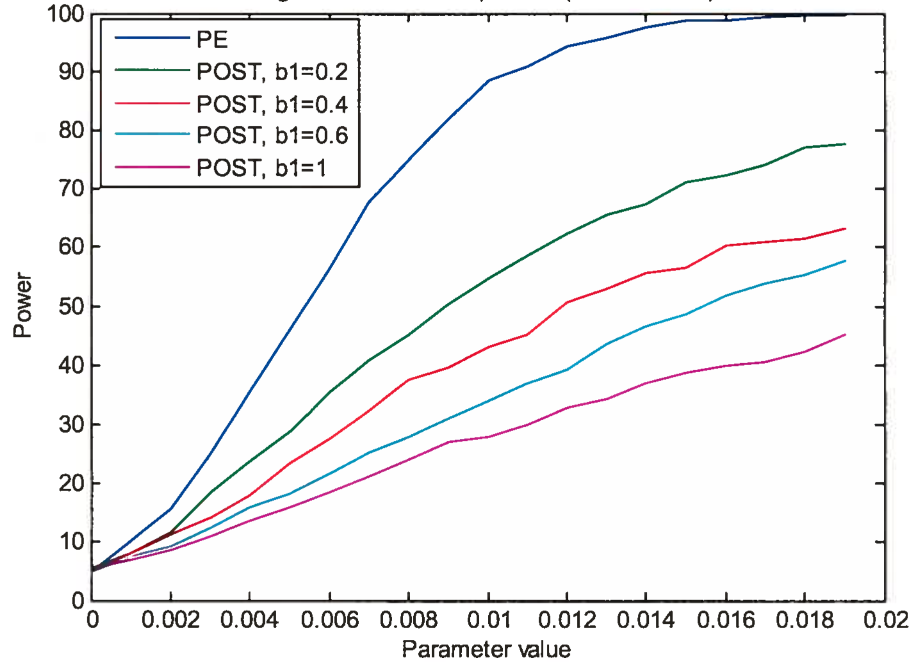
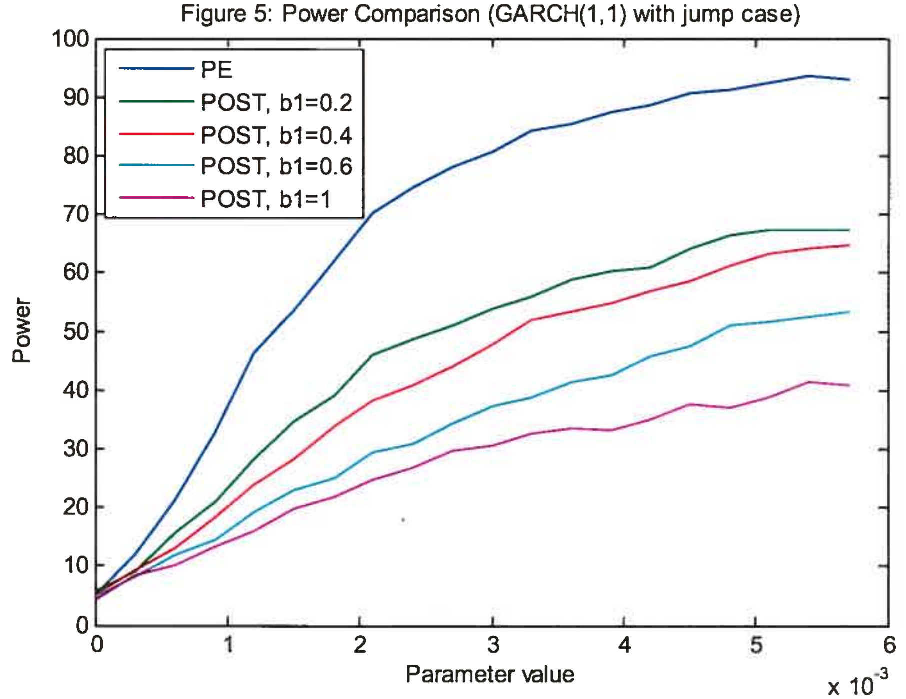


Figure 5: Power Comparison (GARCH(1,1) with jump case)



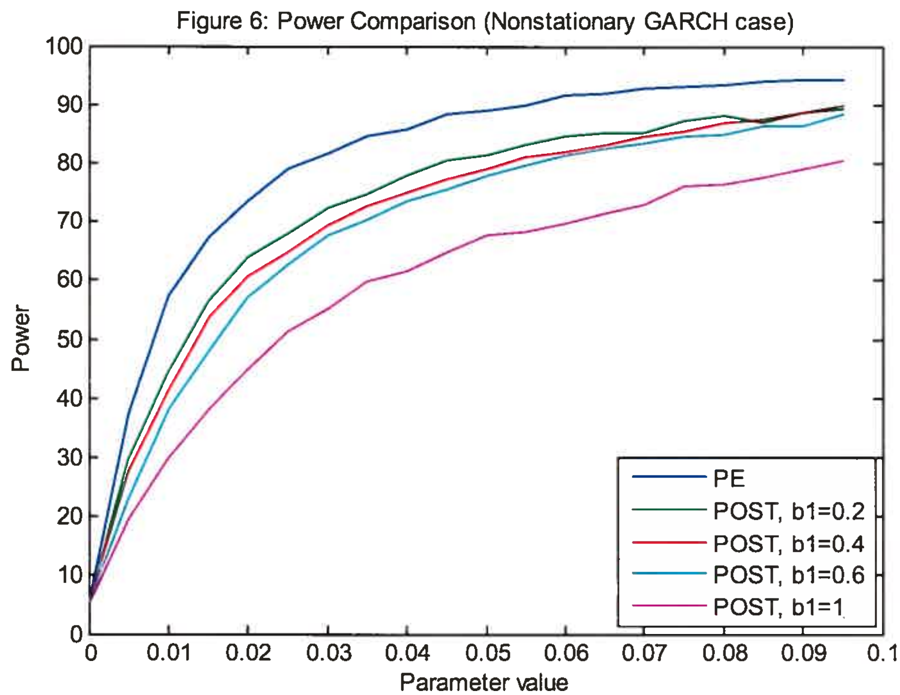


Figure 6: Power Comparison (Nonstationary GARCH case)
Power
Parameter value
PE
POST, b1=0.2
POST, b1=0.4
POST, b1=0.6
POST, b1=1

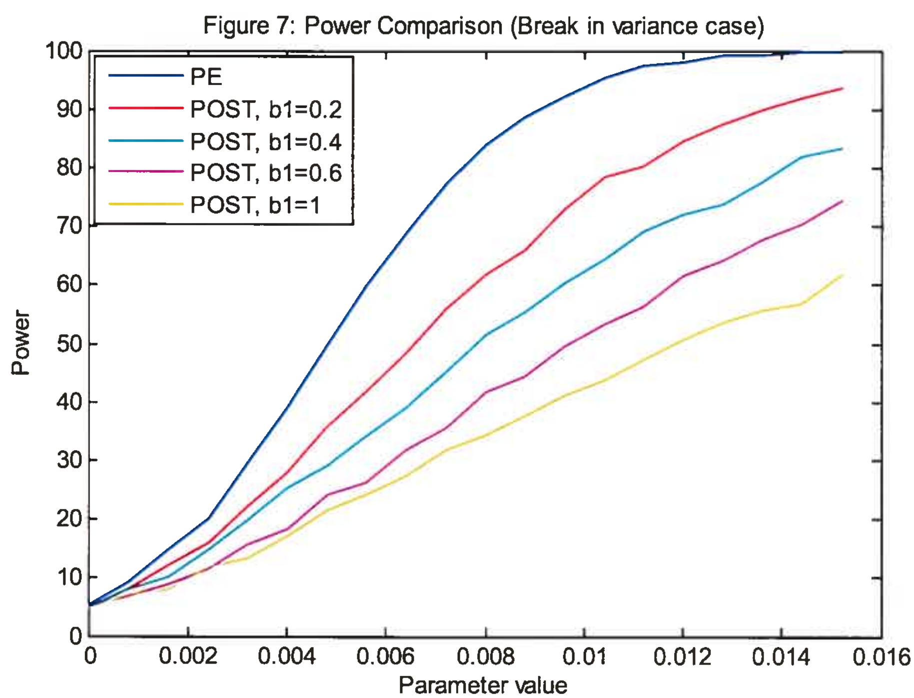


Figure 7: Power Comparison (Break in variance case)
Power
Parameter value
PE
POST, b1=0.2
POST, b1=0.4
POST, b1=0.6
POST, b1=1

Figure 8: Power Comparison (Normal case)

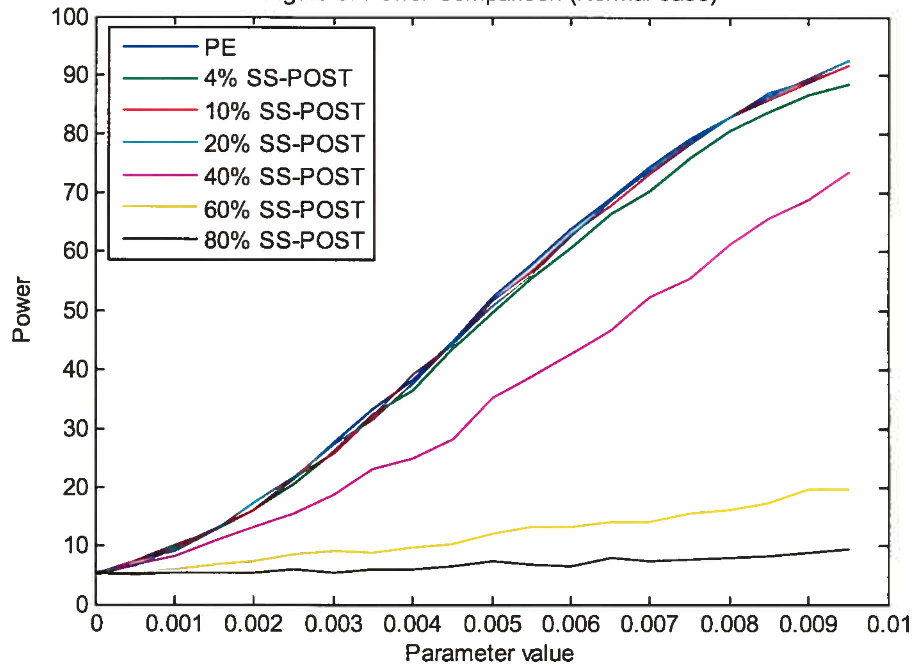


Figure 9: Power Comparison (Cauchy case)

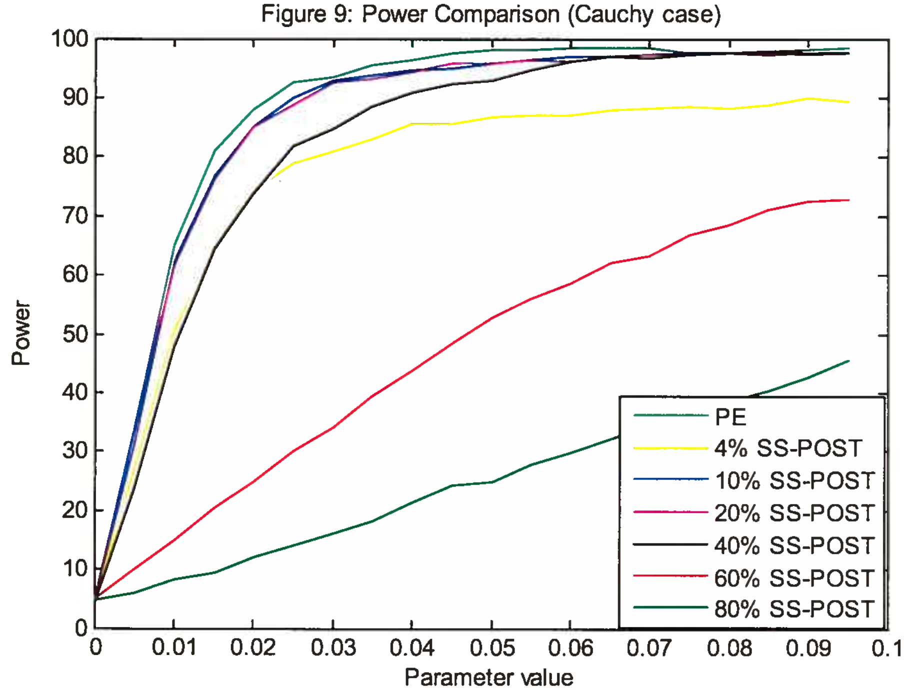


Figure 10: Power Comparison (Mixture case)

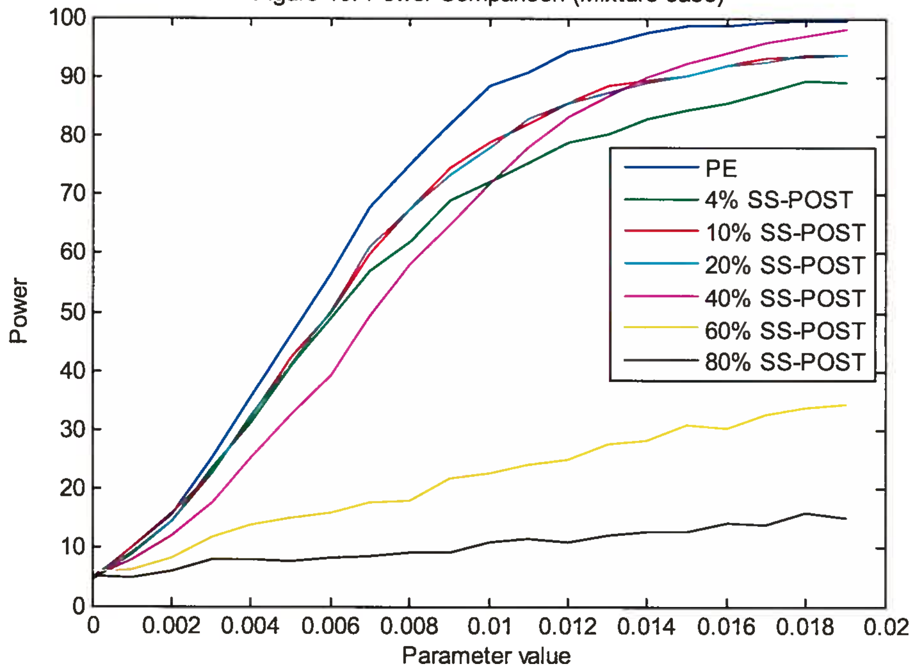
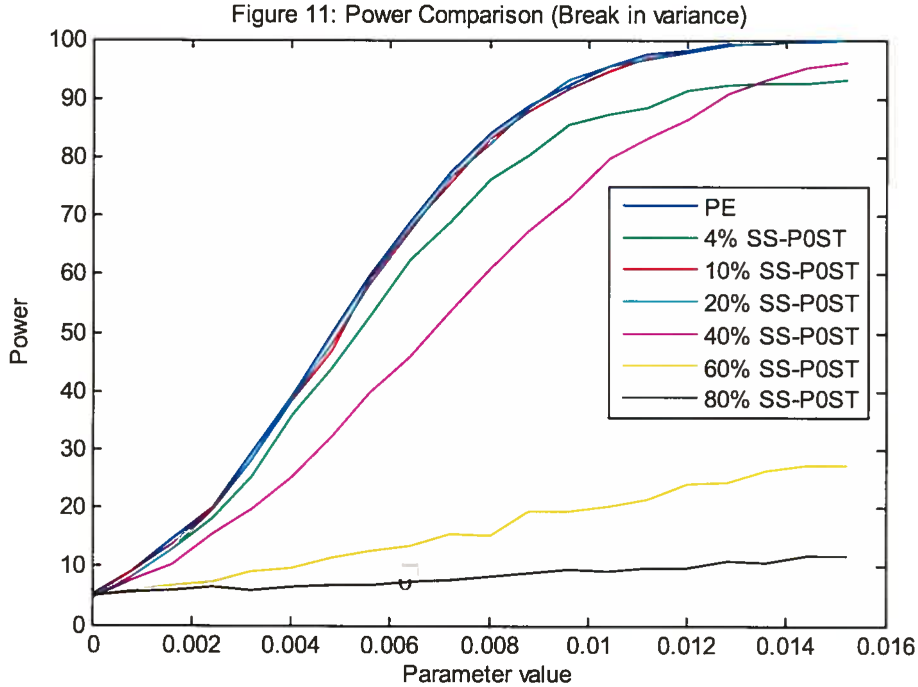


Figure 11: Power Comparison (Break in variance)



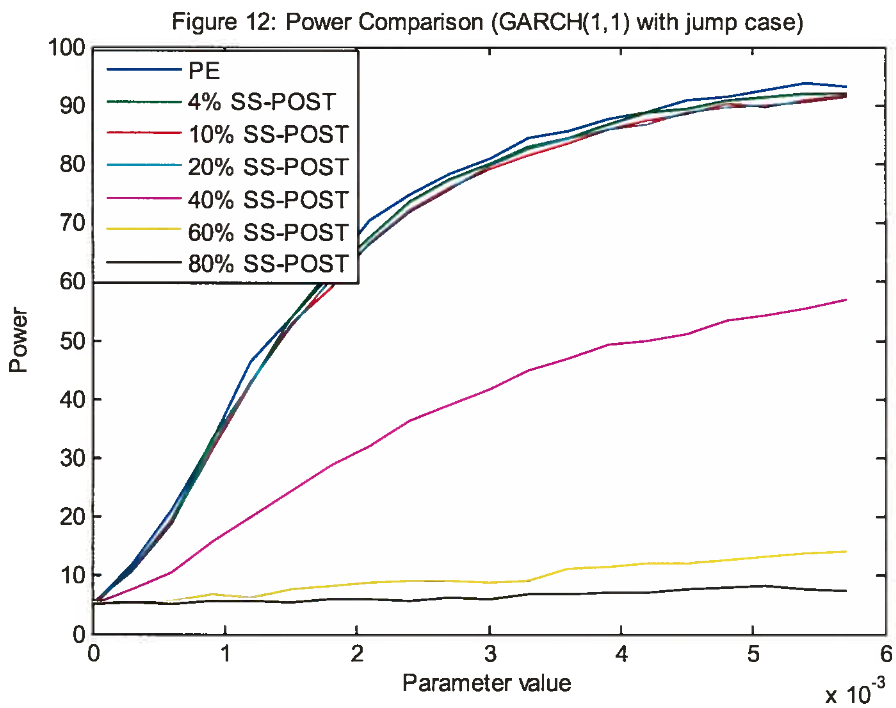


Figure 12: Power Comparison (GARCH(1,1) with jump case)
PE
4% SS-POST
10% SS-POST
20% SS-POST
40% SS-POST
60% SS-POST
80% SS-POST
Power
Parameter value
x 10^-3
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6

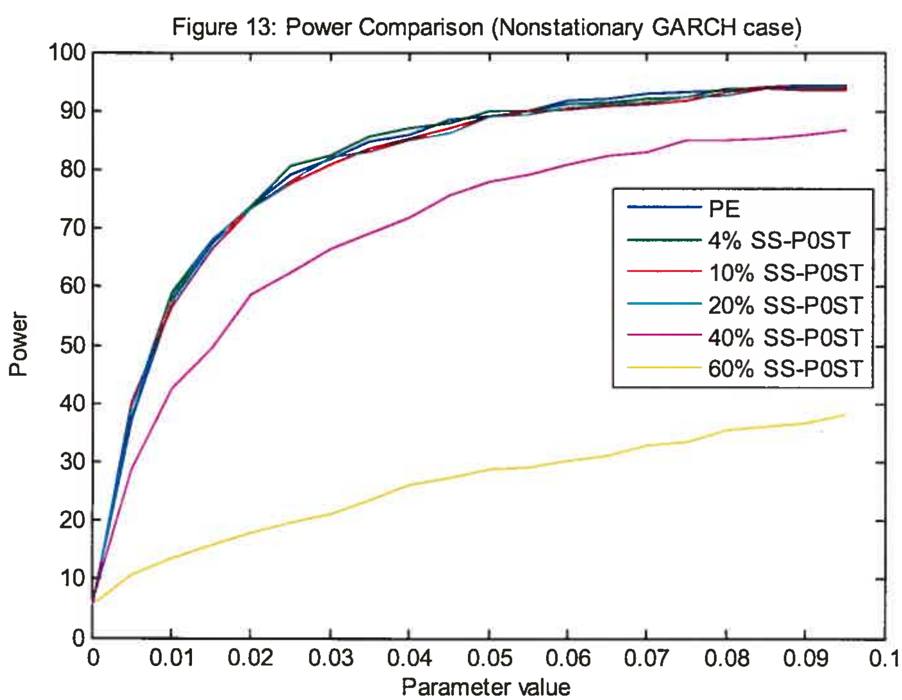


Figure 13: Power Comparison (Nonstationary GARCH case)
PE
4% SS-P0ST
10% SS-P0ST
20% SS-P0ST
40% SS-P0ST
60% SS-P0ST
Power
Parameter value
100
90
80
70
60
50
40
30
20
10
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1

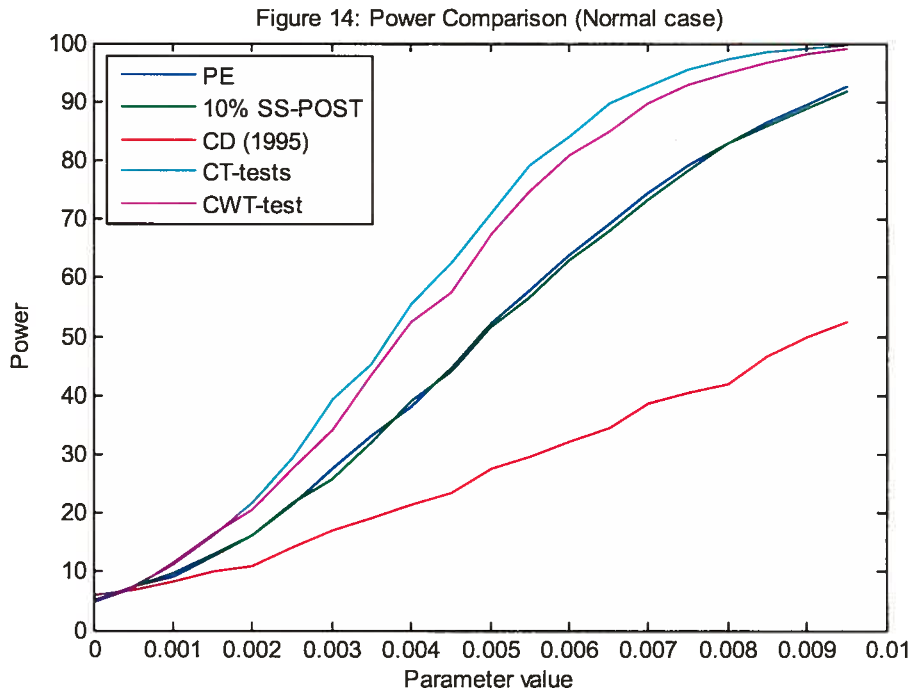


Figure 14: Power Comparison (Normal case)
PE
10% SS-POST
CD (1995)
CT-tests
CWT-test
Power
Parameter value
0
10
20
30
40
50
60
70
80
90
100
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008
0.009
0.01

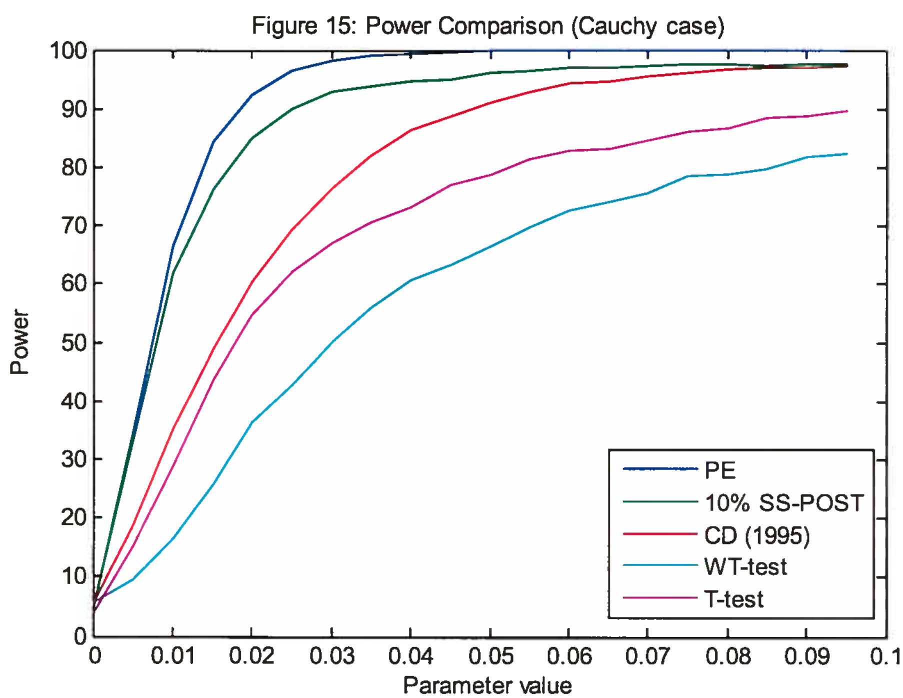


Figure 15: Power Comparison (Cauchy case)
PE
10% SS-POST
CD (1995)
WT-test
T-test
Power
Parameter value
0
10
20
30
40
50
60
70
80
90
100
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1

Figure 16: Power Comparison (Mixture case)

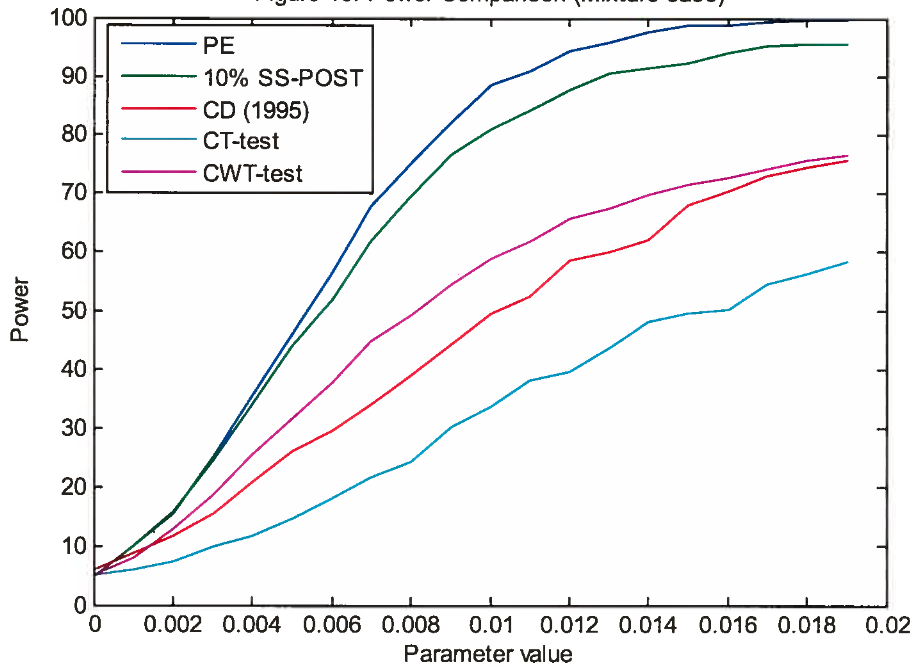
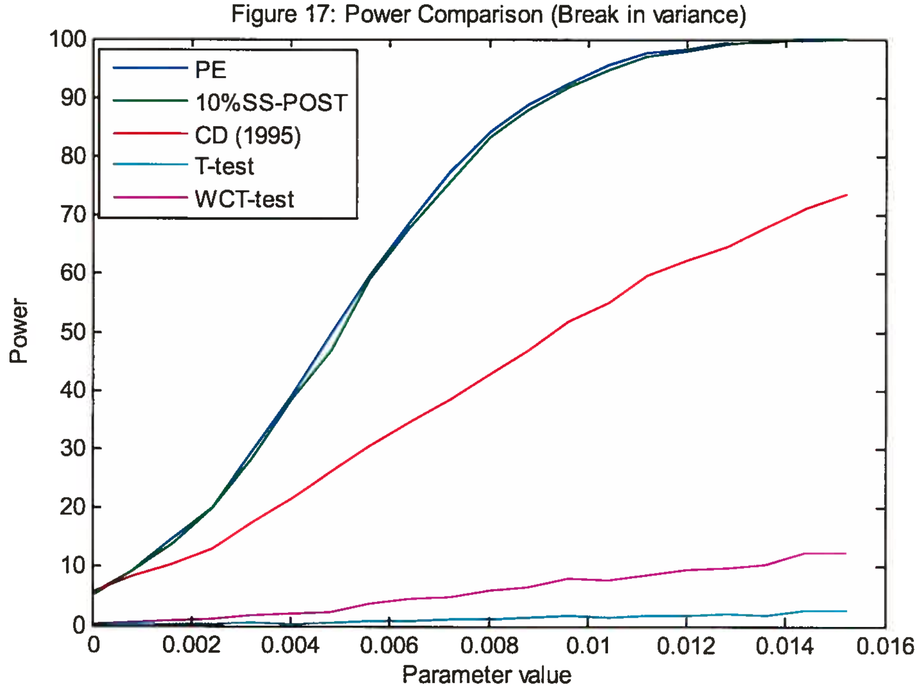


Figure 17: Power Comparison (Break in variance)



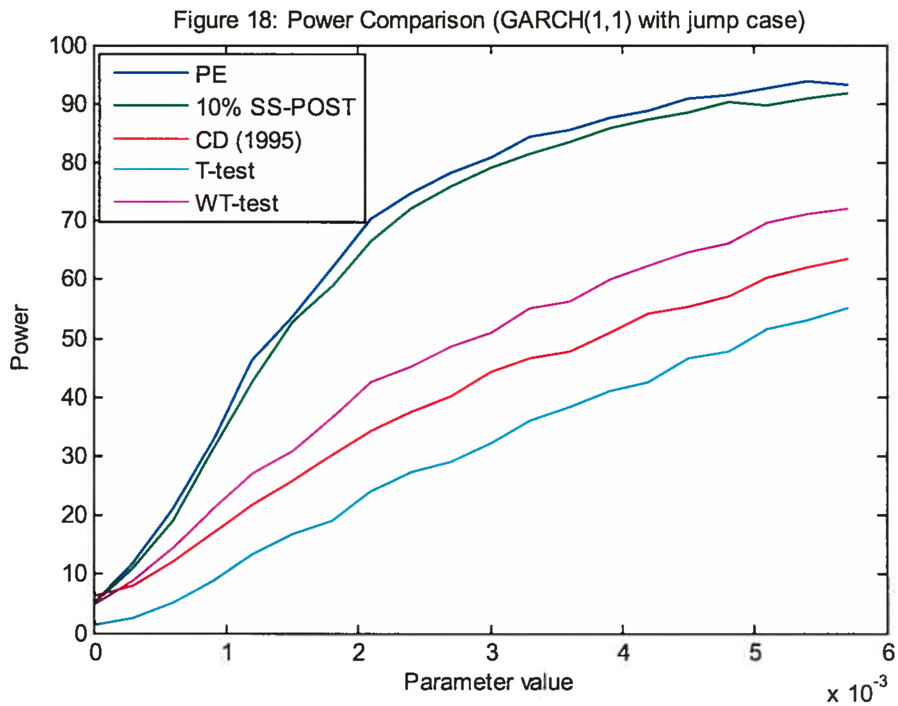


Figure 18: Power Comparison (GARCH(1,1) with jump case)
PE
10% SS-POST
CD (1995)
T-test
WT-test
Power
Parameter value
x 10^-3

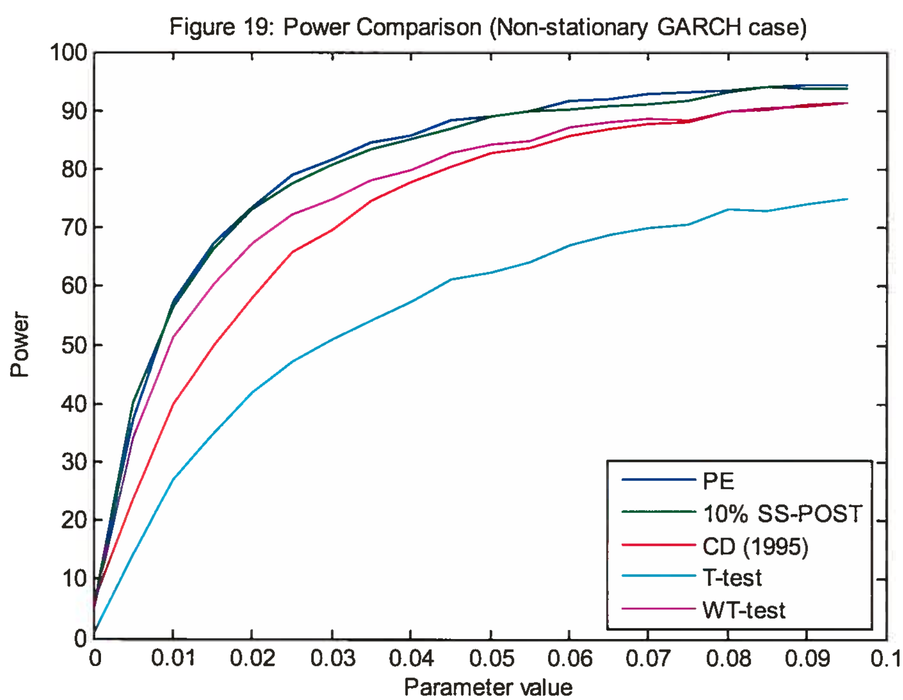


Figure 19: Power Comparison (Non-stationary GARCH case)
PE
10% SS-POST
CD (1995)
T-test
WT-test
Power
Parameter value

Figure 20: Power Comparison (Non Stationary GARCH(1,1) case)

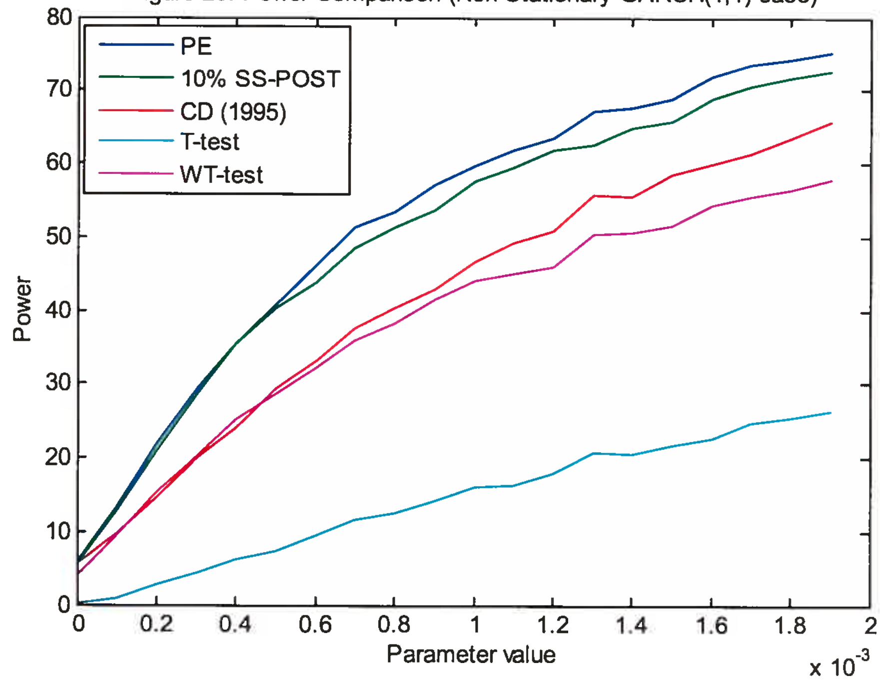
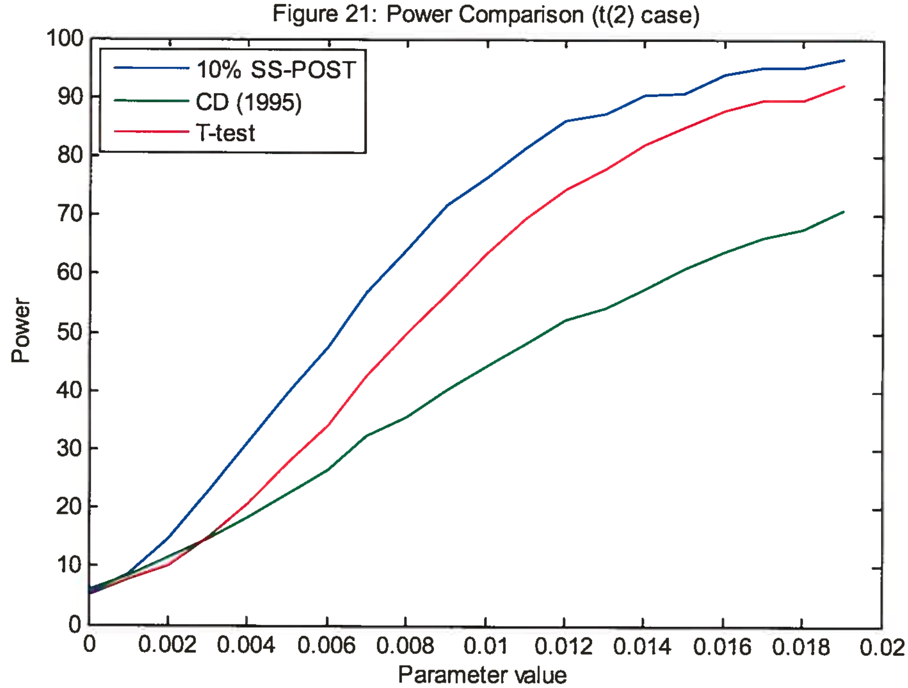


Figure 21: Power Comparison (t(2) case)



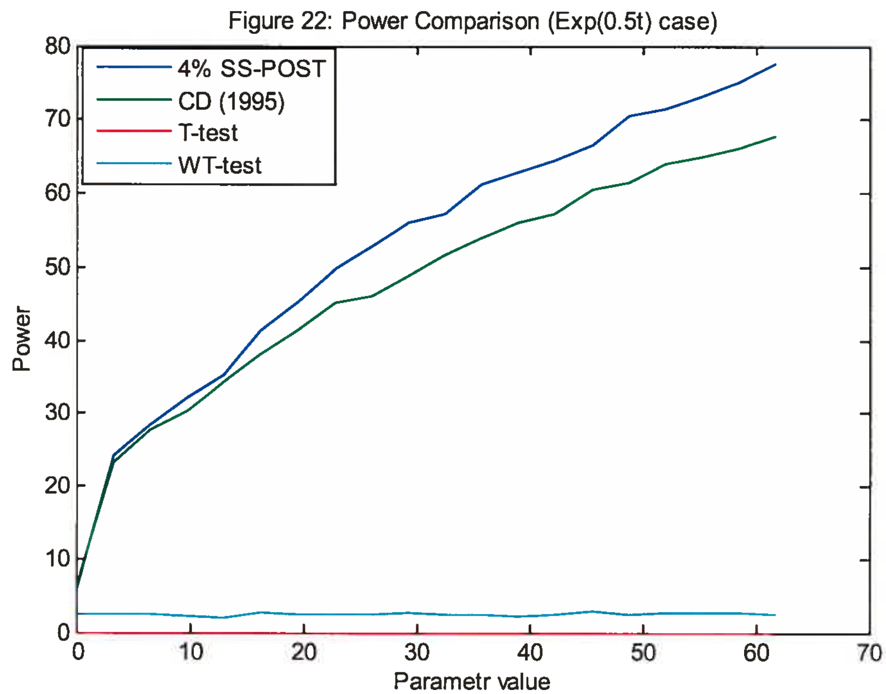


Figure 22: Power Comparison (Exp(0.5t) case)
4% SS-POST
CD (1995)
T-test
WT-test
Power
Parametr value
0
10
20
30
40
50
60
70
80

		POST		Split − Sample				Other − Tests			
β	PE	$\beta_1 = 0.2$	$\beta_1 = 0.4$	4%	10%	20%	40%	CD (1995)	$T - test$	$WT - test$	$CWT - test$
0	5.2	5.14	5.34	4.82	4.88	5.36	4.78	5.94	4.88	7.52	4.94
0.0005	7.44	5.96	6.5	7.58	7.44	6.62	6.78	6.96	7.42	10.7	7.3
0.001	9.2	8.24	7.96	9.98	9.82	9.48	8.2	8.24	11.4	15.4	11.5
0.0015	12.78	11.28	10.24	12.6	12.9	12.76	11.04	10.06	16.24	20.08	16.5
0.002	16.34	13.34	11.96	16.28	16.18	17.26	13.18	11.02	21.7	26.78	20.68
0.0025	21.38	16.36	14.02	20.56	21.8	21.7	15.76	14.12	29.42	34.42	27.74
0.003	27.74	20.74	17.62	26.08	25.84	27.26	18.74	17.02	39.32	41.2	34.24
0.0035	33.26	23.48	20.86	32.44	32.08	31.42	23.28	19.22	45.22	49.16	43.48
0.004	38.14	28.28	23.46	36.4	39.08	37.52	24.88	21.56	55.36	58.52	52.38
0.0045	44.68	32.68	27.68	43.28	44.1	44.3	28.14	23.46	62.38	66.96	57.44
0.005	52.2	36.68	29.7	49.44	51.74	50.6	35.24	27.5	71.04	73.16	67.32
0.0055	57.76	40.78	33.5	55.42	56.68	56.06	38.64	29.8	79.16	79.92	74.7
0.006	63.92	45.44	37.26	60.78	63.12	62.62	42.44	32.3	84.18	85.7	80.84
0.0065	69.22	47.66	40.68	66.44	68	68.9	46.74	34.78	89.58	89.74	85.06

Table 1: Power comparison: Normal case

PE=Power envelope.

$POST$= Point-optimal sign test.

$Split - Sample$ =Point-optimal sign test based on sample split technique.

CD(1995)= Campbell and Dufour's (1995) test.

$WT - test$=$T - test$ based on White's (1980) correction of variance.

$CWT - test$=$WT - test$ with correction of size.

		$POST$		$Split-Sample$				$Other-Tests$		
β	PE	$\beta_1=0.2$	$\beta_1=0.4$	4%	10%	20%	40%	CD (1995)	$T-test$	$WT-test$
0	5.1	4.88	4.8	5.02	5.3	5.48	4.46	5.78	5.68	3.94
0.005	34.22	25.18	20.94	26.72	33.3	30.86	23.48	18.44	9.5	15
0.01	66.38	48.42	39.58	50.46	61.74	62.28	47.86	35.16	16.6	28.92
0.015	84.44	62.56	52.94	64.74	76.24	77.02	64.38	48.9	25.76	43.82
0.02	92.2	74.3	63.08	74.36	84.9	85.14	73.7	60.36	36.28	54.72
0.025	96.44	79.62	69.6	79.06	89.88	88.82	81.78	69.58	42.74	62.08
0.03	98.12	82.86	74.3	81.08	92.92	92.58	84.7	76.6	50.14	67.06
0.035	99	86.02	78.36	82.86	93.7	93.1	88.38	81.88	56	70.72
0.04	99.36	89.16	79.6	85.62	94.7	94.3	90.76	86.42	60.56	73.34
0.045	99.68	89.92	81.88	85.74	94.92	95.74	92.24	88.84	63.3	77.18
0.05	99.8	91.12	84.24	86.76	95.92	95.92	93	91.18	66.6	78.7
0.055	99.98	91.94	86.2	87.14	96.42	96.48	94.56	92.98	69.88	81.3
0.06	99.94	92.5	86.38	87.08	97.02	96.18	95.96	94.16	72.72	82.96
0.065	99.94	93.08	86.84	88.02	96.86	96.9	96.92	94.68	74.1	83.22

Table 2: Power comparison: Cauchy case

PE=Power envelope.

$POST$= Point-optimal sign test.

$Split-Sample$ =Point-optimal sign test based on sample split technique.

CD(1995)= Campbell and Dufour's (1995) test.

$WT-test$=T-test based on White's (1980) correction of variance.

		POST		Split – Sample				Other – Tests				
β	PE	$\beta_1 = 0.2$	$\beta_1 = 0.4$	4%	10%	20%	40%	CD	$Ttest$	$WTtest$	$CTtest$	$CWTtest$
0	4.96	5.3	4.9	4.58	4.7	5.02	5.18	5.98	9.92	10.74	5.08	5.04
0.001	9.96	8.08	8.14	8.86	9.98	9.16	8.02	8.94	11.28	13.12	5.9	7.92
0.002	15.7	11.52	11.3	14.46	15.9	14.6	12.24	11.76	13.98	18.88	7.5	12.94
0.003	25.26	18.48	14.24	23.5	26.26	26.1	21.14	15.72	16.9	25.76	10.1	18.74
0.004	35.46	23.84	18.12	31.1	35.58	35.78	28.86	21	20.68	31.76	11.82	25.68
0.005	46.08	28.7	23.66	40.66	45.64	44.46	36.1	26.24	24.32	40.04	14.64	31.74
0.006	56.68	35.52	27.56	48.94	53.28	53.56	42.72	29.72	28.24	47.06	18.16	37.82
0.007	67.64	40.66	32.3	56.84	63.4	64.44	52.66	34.06	33	51.22	21.92	44.76
0.008	75	45.32	37.46	61.94	70.98	71	61.6	38.96	36.62	56.7	24.56	49.14
0.009	82.06	50.4	39.64	68.78	78.02	76.82	68.18	44.22	40.16	60.5	30.18	54.6
0.01	88.48	54.9	43.24	72.2	82.34	81.4	75.18	49.58	45.86	63.74	33.64	58.8
0.011	90.68	58.48	45.24	75.42	85.66	86.44	81.42	52.4	48.6	66.9	38.06	61.7
0.012	94.38	62.44	50.78	78.94	89.16	88.92	86.68	58.54	51.16	69.26	39.72	65.62
0.013	95.7	65.76	53.12	80.32	92.04	90.72	90.14	60.1	55.26	72.16	43.66	67.42

Table 3: Power comparison: Mixture case

PE=Power envelope.

$POST$= Point-optimal sign test

$Split - Sample$ =Point-optimal sign test based on sample split technique

CD= Campbell and Dufour's (1995) test.

$WTtest$=$Ttest$ based on White's variance correction.

$CTtest = Ttest$ with size correction.

$CWTtest$=$WT - test$ with size correction.

		POST		Split − Sample				Other − Tests		
β	PE	$\beta_1 = 0.2$	$\beta_1 = 0.4$	4%	10%	20%	40%	$CD(1995)$	$T - test$	$WT - test$
0	5.4	4.98	4.92	4.84	5.24	5.1	4.96	5.78	0.01	0.16
0.0008	9.22	7.96	7.9	8.28	9.32	8.38	7.68	8.24	0.04	0.42
0.0016	14.78	12	10.18	13.12	13.76	12.98	10.42	10.44	0.06	0.6
0.0024	20.16	15.88	14.62	18.2	20.12	19.86	15.58	12.98	0.12	1.08
0.0032	29.32	22.12	19.6	25.24	28.34	28.26	19.64	17.34	0.3	1.62
0.004	39.04	27.96	25.38	35.72	38.32	38.68	25.24	21.4	0.22	1.86
0.0048	49.78	35.7	29.12	43.98	47	48.06	32.38	26.12	0.46	2.3
0.0056	59.66	41.62	34.12	52.82	59.16	58.24	39.78	30.42	0.84	3.6
0.0064	68.88	48.5	39.14	62.3	67.9	67.28	45.96	34.78	0.78	4.58
0.0072	77.32	55.9	45.3	68.78	75.66	76.5	53.54	38.38	0.94	4.88
0.008	83.96	61.9	51.68	76.14	83.14	82.2	60.92	42.72	0.94	5.88
0.0088	88.76	65.9	55.52	80.14	88	88.5	67.46	47.04	1.22	6.54
0.0096	92.22	72.94	60.32	85.6	91.7	93.02	73.06	51.76	1.5	8.14
0.0104	95.42	78.52	64.48	87.42	94.68	95.34	79.76	55.02	1.42	7.88

Table 4: Power comparison: Break in variance case

PE=Power envelope.

$POST$= Point-optimal sign test.

$Split - Sample$ =Point-optimal sign test based on sample split technique.

CD(1995)= Campbell and Dufour's (1995) test.

$WT - test$=T-test based on White's (1980) correction of variance.

		$POST$		$Split-Sample$				$Other-Tests$		
β	PE	$\beta_1 = 0.2$	$\beta_1 = 0.4$	4%	10%	20%	40%	CD (1995)	$T-test$	$WT-test$
0	5.07	5.74	4.98	4.7	5.24	5.4	5.04	6.42	1.22	4.96
0.0003	11.98	9.06	9.16	11.18	11.02	10.76	7.86	8.06	2.36	8.92
0.0006	21.28	15.5	12.9	19.38	19.2	18.84	10.74	12.18	5	14.6
0.0009	32.8	21	18.14	33.12	31.34	32.12	15.98	17.24	8.9	21.2
0.0012	46.28	28.14	23.9	42.46	42.46	42.72	19.98	21.9	13.36	27.16
0.0015	53.62	34.62	28.2	53.52	52.7	52.2	24.56	25.86	16.76	30.86
0.0018	62.24	39.1	33.74	61.36	59	60.4	28.8	30.12	19.06	36.58
0.0021	70.22	46.06	38.1	67.52	66.44	66.14	31.96	34.44	24.2	42.58
0.0024	74.66	48.74	40.72	73.66	71.94	71.8	36.28	37.68	27.26	45.1
0.0027	78.28	50.88	43.94	77.36	75.98	75.44	38.98	40.12	29.22	48.82
0.003	80.72	54.04	47.76	79.96	79.22	79.66	41.54	44.32	32.4	51.02
0.0033	84.22	56.12	51.8	82.76	81.38	82.62	44.96	46.72	36.1	55.08
0.0036	85.42	58.82	53.44	84.46	83.52	84.5	47	47.84	38.32	56.42
0.0039	87.66	60.52	54.78	86.58	85.76	85.94	49.18	51.04	41.22	60.18

Table 5: Power comparison: Outlier in $GARCH(1,1)$ case

PE=Power envelope.

$POST$= Point-optimal sign test.

$Split-Sample$ =Point-optimal sign test based on sample split technique.

$CD(1995)$= Campbell and Dufour's (1995) test.

$WT-test$=T-test based on White's (1980) correction of variance.

		$POST$		$Split - Sample$				$Other - Tests$		
β	PE	$\beta_1 = 0.2$	$\beta_1 = 0.4$	4%	10%	20%	40%	CD (1995)	$T - test$	$WT - test$
0	5.95	5.58	6.08	6.02	5.76	6.04	6.16	6.26	0.94	5
0.005	37.34	29.68	27.72	39.04	40.28	39	28.78	23.58	14.26	34.18
0.01	57.36	44.54	41.36	58.86	56.58	58.04	42.64	39.78	27	51.22
0.015	67.3	56.54	53.58	67.92	66.54	68	49.7	49.84	35	60.44
0.02	73.46	63.76	60.56	73.64	73.16	73.36	58.74	58.04	42.04	67.28
0.025	79.02	67.86	64.7	80.6	77.64	78.04	62.34	65.88	47.16	72.36
0.03	81.66	72.5	69.38	82.18	80.88	81.88	66.6	69.72	50.9	75.14
0.035	84.58	74.72	72.56	85.4	83.42	82.8	69.18	74.78	54.22	78.24
0.04	85.82	77.86	75.08	86.86	85.3	84.82	71.84	77.82	57.52	80.04
0.045	88.46	80.52	77.2	87.98	86.9	86.12	75.46	80.44	61.18	82.96
0.05	89.02	81.48	79.22	89.92	89.1	88.98	77.84	83.04	62.48	84.34
0.055	90.04	83.2	81	89.94	89.94	89.22	79.08	83.82	64.16	84.88
0.06	91.76	84.52	81.96	91.14	90.1	90.5	80.86	85.7	67.2	87.26
0.065	91.82	85.22	83.22	91.3	90.86	91.12	82.38	87	68.8	88.22

Table 6: Power comparison: Non-stationary $GARCH$ case

PE=Power envelope.

$POST$= Point-optimal sign test

$Split - Sample$ =Point-optimal sign test based on sample split technique

CD(1995)= Campbell and Dufour's (1995) test.

$WT - test$=T-test based on White's variance correction.

Bibliography of Chapter 1

Berkowitz, J. and L. Kilian. (2000). "Recent Developments in Bootstrapping Time Series," Econometric Reviews, 19, 1–48.

Bernanke, B. S. and I. Mihov. (1998). "Measuring Monetary Policy," The Quarterly Journal of Economics, 113(3), 869–902.

Bhansali, R. J. (1978). "Linear Prediction by Autoregressive Model Fitting in the Time Domain," Ann. Statist., 6, 224–231.

Boudjellaba, H., J.-M. Dufour, and R. Roy. (1992). "Testing Causality Between Two Vectors in Multivariate ARMA Models," Journal of the American Statistical Association, 87, 1082–1090.

Boudjellaba, H., J.-M. Dufour, and R. Roy. (1994). "Simplified Conditions for Non-Causality between Two Vectors in Multivariate ARMA Models," Journal of Econometrics, 63, 271–287.

Diebold, F. X. and L. Kilian. (2001). "Measuring Predictability. Theory and Macroeconomic Applications," Journal of Applied Econometrics, 16, 657–669.

Dufour, J.-M. and D. Pelletier. (2005). "Practical methods for modelling weak VARMA processes: Identification, estimation and specification with a macroeconomic application," Technical report, Département de sciences économiques and CIREQ, Université de Montréal, Montréal, Canada.

Dufour J.-M. and E. M. Iglesias. (2007). "Finite sample and optimal adaptive inference in possibly nonstationary general volatility models with Gaussian or heavy-tailed errors," Mimeo.

Dufour, J.-M., D. Pelletier, and É. Renault. (2006). "Short Run and Long Run Causality in Time Series: Inference," Journal of Econometrics 132 (2), 337–362.

Dufour, J.-M. and T. Jouini. (2004). "Asymptotic Distribution of a Simple Linear Estimator for VARMA Models in Echelon Form," forthcoming in Statistical Modeling and Analysis for Complex Data Problems, ed. by Pierre Duchesne and Bruno Remillard, Kluwer, The Netherlands.

Dufour, J.-M. and E. Renault. (1998). "Short-Run and Long-Run Causality in Time Series. Theory," Econometrica, 66(5), 1099–1125.

Efron, B. and R. J. Tibshirani. (1993). "An Introduction to the Bootstrap," New York. Chapman Hall.

Geweke, J. (1982). "Measurement of Linear Dependence and Feedback between Multiple

Time Series," Journal of the American Statistical Association, 77(378), 304–313.

Geweke, J. (1984). "Measures of Conditional Linear Dependence and Feedback between Time Series," Journal of the American Statistical Association, 79, 907–915.

Geweke, J. (1984a). "Inference and Causality in Economic Time Series," in Handbook of Econometrics, Volume 2, ed. by Z. Griliches and M. D. Intrilligator. Amsterdam. North-Holland, pp. 1102-1144.

Gouriéroux, C., A. Monfort, and E. Renault. (1987). "Kullback Causality Measures," Annales D'Économie et De Statistique, 6/7, 369–410.

Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross- Spectral Methods," Econometrica, 37, 424–459.

Hannan, E. J., and L. Kavalieris. (1984b). "Multivariate Linear Time Series Models," Advances in Applied Probability, 16, 492–561.

Hannan, E. J., and J. Rissanen. (1982). "Recursive Estimation of Mixed Autoregressive-Moving Average Order," Biometrika, 69, 81–94. Errata 70 (1983), 303.

Hsiao, C. (1982). "Autoregressive Modeling and Causal Ordering of Economic Variables," Journal of Economic Dynamics and Control, 4, 243–259.

Inoue, A. and L. Kilian. (2002). "Bootstraping Smooth Functions of Slope Parameters and Innovation Variances in VAR(Infinite) Models," International Economic Review, 43, 309–332.

Kang, H. (1981). "Necessary and Sufficient Conditions for Causality Testing in Multivariate ARMA Model," Journal of Time Series Analysis, 2, 95–101.

Kilian, L. (1998). "Small-sample confidence intervals for impulse response functions," Review of Economics and Statistics 80, 218–230.

Koreisha, S. G. and T. M. Pukkila. (1989). "Fast Linear Estimation Methods for Vector Autorgessive Moving-Average Models," Journal of Time Series Analysis, 10(4), 325–339.

Lewis, R. and G. C. Reinsel. (1985). "Prediction of Multivariate Time Series by Autoregressive Model Fitting," Journal of Multivariate Analysis, 16, 393–411.

Lütkepohl, H. (1993a). "Introduction to Multiple Time Series Analysis," second edn, Springer-Verlag, Berlin.

Lütkepohl, H. (1993b). "Testing for causation between two variables in higher dimensional VAR models," in H. Schneeweiss and K. Zimmermann, eds, 'Studies in Applied Econometrics', Springer-Verlag, Heidelberg.

Newbold, P. (1982). "Causality Testing in Economics. in Time Series Analysis. Theory and Practice 1," ed. by O. D. Anderson. Amsterdam. North-Holland.

Paparoditis, E. (1996). ''Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes,'' Journal of Multivariate Analysis 57, 277–96.

Parzen, E. (1974). "Some Recent Advances in Time Series Modelling," IEEE Trans. Automat. Control, AC-19.

Patterson, K. (2007). "Bias Reduction Through First-Order Mean Correction, Bootstrapping and Recursive Mean Adjustment," Journal of Applied Statistics, 34, 23–45.

Pierce, D. A. and L. D. Haugh. (1977). "Causality in Temporal Systems. Characterizations and Survey," Journal of Econometrics, 5, 265–293.

Polasek, W. (1994). "Temporal Causality Measures Based on AIC," in H. Bozdogan, Proceedings of the Frontier of Statistical Modeling. An Informal Approach, Kluwer, Netherlands, pp. 159– 168.

Polasek, W. (2000). "Bayesian Causality Measures for Multiple ARCH Models Using Marginal Likelihoods," Working Paper.

Sims, C. (1972). "Money, Income and Causality," American Economic Review, pp. 540–552.

Sims, C. (1980). "Macroeconomics and Reality," Econometrica, 48, 1–48.

Wiener, N. (1956). "The Theory of Prediction," In The Theory of Prediction, Ed by E. F. Beckenback. New York. McGraw-Hill, Chapter 8.

Bibliography of Chapter 2

Andersen, T. and Sorensen, B. (1994). "Estimation of a Stochastic Volatility Model: A Monte Carlo Study," Journal of Business and Economic Statistics 14, 328–352.

Andersen, T.G., T. Bollerslev, and F.X. Diebold. (2003). "Some Like it Smooth, and Some Like it Rough. Untangling Continuous and Jump Components in Measuring, Modeling, and Forecasting Asset Return Volatility," Working Paper.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens. (2001). "The Distribution of

Stock Return Volatility," Journal of Financial Economics, 61, 1, 43-76.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. (2001). "The Distribution of Realized Exchange Rate Volatility," Journal of the American Statistical Association, 96, 42-55.

Andersen, T.G., T. Bollerslev, and F.X. Diebold. (2003). "Parametric and Non-Parametric Volatility Measurement," Handbook of Financial Econometrics (L.P Hansen and Y. Aït-Sahalia, eds.). Elsevier Science, New York, forthcoming.

Andersen, T.G. and T. Bollerslev. (1998). "Answering the Skeptics. Yes, Standard Volatility Models Do Provide Accurate Forecasts," International Economic Review, 39, 885-905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega. (2003). "Micro Effects of Macro Announcements. Real-Time Price Discovery in Foreign Exchange," American Economic Review, 93, 38-62.

Ang, A. and J. Liu. (2006). "Risk, Return, and Dividends," forthcoming Journal of Financial Economics.

Balduzzi, P., E. J. Elton, and T. C. Green. (2001). "Economic News and Bond Prices. Evidence from the U.S. Treasury Market," Journal of Financial and Quantitative Analysis, 36, 523-544.

Barndorff-Nielsen, O.E. and N. Shephard. (2002a). "Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models," Journal of the Royal Statistical Society, 64, 253-280.

Barndorff-Nielsen, O.E. and N. Shephard. (2002b). "Estimating Quadratic Variation Using Realized Variance," Journal of Applied Econometrics, 17, 457-478.

Barndorff-Nielsen, O.E. and N. Shephard. (2003c). "Power and Bipower Variation with Stochastic and Jumps," Manuscript, Oxford University.

Barndorff-Nielsen, O.E., S.E. Graversen, J. Jacod, M. Podolskij, and N. Shephard. (2005). "A Central Limit Theorem for Realized Power and Bipower Variations of Continuous Semimartingales." Working Paper, Nuffield College, Oxford University; forthcoming in Yu Kabanov and Robert Liptser (eds.), From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev. New York. Springer-Verlag.

Bekaert, G. and G. Wu. (2000). "Asymmetric Volatility and Risk in Equity Markets," The Review of Financial Studies, 13, 1-42.

Black, F. (1976). "Studies of Stock Price Volatility Changes," Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics, 177-

181.

Bollerslev, T. and H. Zhou. (2005). "Volatility Puzzles. A Unified Framework for Gauging Return-Volatility Regressions," Journal of Econometrics, forthcoming.

Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen. (2005). "A Discrete-Time Model for Daily S&P500 Returns and Realized Variations. Jumps and Leverage Effects," Working Paper.

Bollerslev, T., J. Litvinova, and G. Tauchen. (2006). "Leverage and Volatility Feedback Effects in High-Frequency Data," Journal of Financial Econometrics 4 (3), 353-384.

Bouchaud, J-P., A. Matacz, and M. Potters. (2001). "Leverage Effect in Financial Markets. The Retarded Volatility Model," Physical Review Letters, 87, 228 701.

Brandt, M. W., and Q. Kang. (2004). "On the Relationship Between the Conditional Mean and Volatility of Stock Returns. A Latent VAR Approach," Journal of Financial Economics, 72, 2004, 217-257.

Campbell, J. and L. Hentschel. (1992). "No News is Good News. An Asymmetric Model of Changing Volatility in Stock Returns," Journal of Financial Economics, 31, 281-331.

Christie, A. C. (1982). "The Stochastic Behavior of Common Stock Variances- Value, Leverage and Interest Rate Effects," Journal of Financial Economics, 3, 145-166.

Comte, F. and E. Renault. (1998). "Long Memory in Continuous Time Stochastic Volatility Models," Mathematical Finance, 8, 291-323.

Corsi, F. (2003). "A Simple Long Memory Model of Realized Volatility," Manuscript, University of Southern Switzerland.

Cutler, D. M., J. M. Poterba, and L. H. Summers. (1989). "What Moves Stock Prices?" The Journal of Portfolio Management, 15, 4-12.

Dacorogna, M.M., R. Gençay, U. Müller, R.B. Olsen, and O.V. Pictet. (2001). "An Introduction to High-Frequency Finance," San Diego. Academic Press.

Dufour J-M. and E. Renault. (1998). "Short-Run and Long-Run Causality in Time Series. Theory," Econometrica 66(5), 1099-1125.

Dufour J-M. and A. Taamouti. (2006). "Nonparametric Short and Long Run Causality Measures," in Proceedings of the 2006 Meetings of the American Statistical Association, Business and Economic Statistics, forthcoming.

Dufour J-M. and A. Taamouti. (2005). "Short and Long Run Causality Measures. Theory and Inference," Working Paper.

Engle, R.F and V.K. Ng. (1993). "Measuring and Testing the Impact of News on Volatility," Journal of Finance, 48, 1749-1778.

French, M., W. Schwert, and R. Stambaugh. (1987). "Expected Stock Returns and Volatility," Journal of Financial Economics, 19, 3-30.

Ghysels, E., P. Santa-Clara and R. Valkanov. (2002). "The MIDAS Touch. Mixed Data Sampling Regression," Discussion Paper UCLA and UNC.

Ghysels, E., P. Santa-Clara, and R. Valkanov. (2004). "There is a risk-return trade-off after all", Journal of Financial Economics, 76, 509-548.

Glosten, L. R., R. Jagannathan, and D. E. Runkle. (1993). "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," Journal of Finance, 48, 1779-1801.

Gouriéroux, C. and A. Monfort. (1992). "Qualitative threshold ARCH models," Journal of Econometrics 52, 159-200.

Granger, C. W. J. (1969). "Investigating causal relations by econometric models and cross-spectral methods," Econometrica 37, 424--459.

Guo, H., and R. Savickas. (2006). "Idiosyncratic Volatility, Stock Market Volatility, and Expected Stock Returns," Journal of Business and Economic Statistics, 24(1), 43-56.

Hardouvelis, G. A. (1987). "Macroeconomic information and stock prices." Journal of Economics and Business, 39, 131-140.

Haugen, A. H., E. Talmor, and W. N. Torous. (1991). "The Effect of Volatility Changes on the Level of Stock Prices and Subsequent Expected Returns," Journal of Finance, 46, 985-1007.

Hull, J. and A. White. (1987). "The pricing of options with stochastic volatilities," Journal of Finance, 42, 281-300.

Huang, X. and G. Tauchen. (2005). "The Relative Contribution of Jumps to Total Price Variance," Working Paper.

Huang, X. (2007). "Macroeconomic News Announcements, Financial Market Volatility and Jumps," Working Paper.

Jacquier, E., N. Polson, and P. Rossi. (2004). "Bayesian Analysis of Stochastic Volatility Models with Leverage Effect and Fat tails," Journal of Econometrics, 122.

Jain, P. C. (1988). "Response of Hourly Stock Prices and Trading Volume to Economic

News," The Journal of Business, 61, 219-231.

Lamoureux, C. G. and G. Zhou. (1996). "Temporary Components of Stock Returns: What Do the Data Tell us?" Review of Financial Studies 9, 1033--1059.

Ludvigson, S. C. and S. Ng. (2005). "The Empirical Risk-Return Relation. A Factor Analysis Approach," Forthcoming, Journal of Financial Economics.

McQueen, G. and V. V. Roley. (1993). "Stock Prices, News, and Business Conditions," The Review of Financial Studies, 6, 683-707.

Meddahi, N. (2002). "A Theoretical Comparison Between Integrated and Realized Volatility," Journal of Applied Econometrics, 17, 475-508.

Müller, U., M. Dacorogna, R. Dav, R. Olsen, O. Pictet, and J. von Weizsacker. (1997). "Volatilities of different time resolutions - analyzing the dynamics of market components," Journal of Empirical Finance 4, 213{39.

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

Pagan, A.R., and G.W. Schwert. (1990). "Alternative models for conditional stock volatility," Journal of Econometrics 45, 267-290.

Pearce, D. K. and V. V. Roley. (1985). "Stock Prices and Economic News," Journal of Business, 58, 49-67.

Pindyck, R.S. (1984). "Risk, Inflation, and the Stock Market," American Economic Review, 74, 334-351.

Schwert, G.W. (1989). "Why Does Stock Market Volatility Change Over Time?" Journal of Finance, 44, 1115-1153.

Schwert, G. W. (1981). "The Adjustment of Stock Prices to Information About Inflation," Journal of Finance, 36, 15-29.

Turner, C.M., R. Startz, and C.R. Nelson. (1989). "A Markov Model of Heteroskedasticity, Risk and Learning in the Stock Market," Journal of Financial Economics, 25, 3-22.

Whitelaw, R. F. (1994). "Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns," The Journal of Finance, 49(2), 515--41.

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

Wu, G. (2001). "The determinants of Asymmetric Volatility," Review of Financial Studies, 14, 837-859.

Yu, J. (2005). ''Is No News Good News? Reconciling Evidence from ARCH and Stochastic Volatility Models,'' Working Paper, Department of Economics, Singapore Management University.

Bibliography of Chapter 3

Balduzzi, P. and A. W. Lynch. (1999). "Transaction Costs and Predictability. Some Utility Cost Calculations," Journal of Financial Economics, 52, 47--78.

Breen, W., L. R. Glosten, and R. Jagannathan. (1989). "Economic Significance of Predictable Variations in Stock Index Returns," Journal of Finance, 44(5), 1177--1189.

Billio, M. and L. Polizzon (2000). "Value at Risk. a Multivariate Switching Regime Model," Journal of Empirical Finance Vol 7, 531-554

Bohmann, H. (1961). "Approximate Fourier analysis of distribution function," Ark. Mat. 4, 99-157.

Bohmann, H. (1970). "A method to calculate the distribution when the characteristic function is known," Nordisk Tidskr. Informationsbehandling (BIT) 10, 237-42.

Bohmann, H. (1972). From Characteristic Function to Distribution function Via Fourier Analysis. Nordisk Tidskr. Informationsbehandling (BIT) 12, 279-83.

Campbell, J. Y. (1987). "Stock Returns and the Term Structure," Journal of Financial Economics, 18, 373--399.

Campbell, J. Y. and R. J. Shiller. (1988). "Stock Prices, Earning, and Expected Dividends," Journal of Finance, 43, 661--676.

Campbell, J. Y., Y. L. Chan, and L. M. Viceira. (2002). "A Multivariate Model of Strategic Asset Allocation," Journal of Financial Economics, forthcoming.

Campbell, J. Y. and L. M. Viceira. (2005). "The Term Structure of the Risk-Return Tradeaff," Working paper.

Cardenas, J., E. Fruchard, E. Koehler, C. Michel, and I. Thomazeau. (1997). "VAR. One Step Beyond," Risk 10 (10), 72-75.

Cooper, M. R. C. Gutierrez, Jr., and W. Marcum. (2001). "On the Predictability of Stock Returns in Real Time," Journal of Business, Forthcoming.

Cooper, M., and H. Gulen. (2001). "Is Time-series Based Predictability Evident in Real-time?," Working Paper.

Davies, R. (1973). " Numerical inversion of a characteristic function," Biometrika 60, 415-417.

Davies, R. (1980). "The distribution of a linear combination of chi-squared random variable," Applied Statistics 29, 323-333.

Duffie, D. and J. Pan. (2001). "Analytical Value-At-Risk with Jumps and Credit Risk," Finance and Stochastics, Vol. 5, No. 2, pp 155-180.

Engle, R. and S. Manganelli. (2002). "CAViaR. Conditional Autoregressive Value At Risk By Regression Quantiles," Forthcoming in Journal of Business and Economic Statistics.

Gil-Pelaez, J. (1951). "Note on the Inversion Theorem, " Biometrika 38, 481-482.

Gomes, F. (2002). "Exploiting Short-Run Predictability," Working Paper, London Business School.

Gordon, J. A., and A. M. Baptista. (2000). "Economic Implications of Using a Mean-VaR Model for Portfolio Selection. A Comparison with Mean-Variance Analysis," Working Paper.

Guidolin, M. and A. Timmermann (2005). "Term Structure of Risk under Alternative Econometric Specifications," Forthcoming in Journal of Econometrics.

Fama, E., and W. Schwert. (1977). "Asset Returns and Inflation," Journal of Financial Economics, 5, 115-146.

Fama, E. F. and K. R. French. (1988). "Dividend Yields and Expected Stock Returns," Journal of Financial Economics, 22, 3--25.

Fama, E. F. and K. R. French. (1989). "Business Conditions and Expected Returns on Stocks and Bonds," Journal of Financial Economics, 25, 23--49.

Feller. W. (1966). "An introduction to probability theory and its applications," Vol.2. New York. Wiley.

Hamilton, D. J. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," Econometrica, Vol 57, No.2, 357-384.

Hamilton, D. J. (1994). "Time Series Analysis," Princeton University Press.

Handa, P. and A. Tiwari. (2004). "Does Stock Return Predictability Imply Improved

Asset Allocation and Performance?," Working Paper, University of Iowa.

Han, Y. (2005). "Can An Investor Profit from Return Predictability in Real Time," Working Paper.

Hodrick, R. J. (1992). "Dividend Yields and Expected Stock Returns. Alternative Procedures for Inference and Measurement," The Review of Financial Studies, 5, 357--386.

Imhof, J. P. (1961). "Computing the Distribution of Quadratic Forms in Normal Variables," Biometrika, 48, 419-426.

Jacobsen, B. (1999). "The Economic Significance of Simple Time Series Models of Stock Return Predictability," Working Paper, University of Amsterdam.

Kandel, S. and R. F. Stambaugh. (1996) . "On the Predictability of Stock Returns. An Asset-Allocation Perspective," Journal of Finance, 51(2), 385--424.

Keim, D. B. and R. F. Stambaugh. (1986). "Predicting Returns in the Stock and Bond Markets," Journal of Financial Economics, 17, 357--390.

Lynch, A. W. (2001). "Portfolio Choice and Equity Characteristics. Characterizing the Hedging Demands Induced by Return Predictability," Journal of Financial Economics, 62, 67--130.

Marquering, W. and M. Verbeek. (2001). "The Economic Value of Predicting Stock Index Returns and Volatility," Working Paper, Tilburg University.

Meddahi, N. and A. Taamouti. (2004). "Moments of Markov Switching Models," Working paper.

Michaud, R. O. (1998). "Efficient Asset Management. A Practical Guide to Stock Portfolio Optimization and Asset Allocation," Harward Business Scholl Press.

Mina, J. and Ulmer, A. (1999). "Delta-Gamma Four ways," Working paper.

Pesaran, M. H. and A. Timmermann. (1995). "Predictability of Stock Returns. Robustness and Economic Significance," Journal of Finance, 50(4), 1201--1228.

RiskMetrics. (1995). Technical Document, JP Morgan. New York, USA.

Rouvinez, C. (1997). "Going Greek with VAR," Risk 10 (2).

Shephard, N. G. (1991,a). "Numerical integration rules for multivariate inversions," Journal of Statistical Computation and Simulation, Vol. 39, pp. 37-46.

Shephard, N. G. (1991,b). "From characteristic function to distribution function. A simple framework for the theory," Economic Theory forthcoming.

Bibliography of Chapter 4

Abdelkhalek, T. and J.-M. Dufour. (1998). "Statistical inference for computable general equilibrium models, with application to a model of the Moroccan economy," Review of Economics and Statistics LXXX, 520.534.

Arrow. K. (1960). "Decision Theory and the Choice of a Level of Significance for the T-Test," In Contributions to Probability and Statistics (Olkin et al., eds.) Stanford University Press, Stanford, California .

Bahadur, R. and L. J. Savage. (1956). "The nonexistence of certain statistical procedures in non-parametric problems," Annals of Mathematical Statistics 27, 1115.22.

Bohmann, H. (1972). "From characteristic function to distribution function via fourier analysis," Nordisk Tidskr. Informationsbehandling (BIT) 12, 279.83.

Boldin, M. V., G. I. Simonova, and Y. N. Tyurin. (1997). "Sign-based methods in linear statistical models," Translations of Mathematical Monographs, American Mathematical Society, Vol. 162.

Campbell, B. and J.-M. Dufour. (1995). "Exact nonparametric orthogonality and random walk tests," Review of Economics and Statistics 77, 1.16.

Campbell, B. and J.-M. Dufour. (1997). "Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter," International Economic Review 38, 151.173.

Christoffersen, P. F. and D. Pelletier. (2004). "Backtesting value-at-risk A duration-based approach," Journal of Financial Econometrics pp. 84.108.

Christoffersen, P. F. (1998). "Evaluating interval forecasts," International Economic Review 39, 841.862.

Coudin, E. and J.-M. Dufour. (2005). "Finite sample distribution-free inference in linear median regressions under heteroskedasticity and nonlinear dependence of unknown form," Technical Report, CREST and Universite de Montreal.

Davies, R. (1973). "Numerical inversion of a characteristic function," Biometrika 60, 415.417.

Davies, R. (1980). "The distribution of a linear combination of chi-squared random variable," Applied Statistics 29, 323.333.

Dufour, J-M. (1997). "Some impossibility theorems in econometrics, with applications to structural and dynamic models," Econometrica 65, 1365.1389.

Dufour, J-M. (2003). "Identification, weak instruments and statistical inference in econometrics," Canadian Journal of Economics 36(4), 767.808.

Dufour, J.-M. and J. Jasiak. (2001). "Finite sample limited information inference methods for structural equations and models with generated regressors," International Economic Review 42, 815.843.

Dufour, J-M. and M. L. King. (1991). "Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors," Journal of Econometrics 47, 115.143.

Dufour, J-M. and J. F. Kiviet. (1998). "Exact inference methods for first-order autoregressive distributed lag models," Econometrica 66, 79.104.

Dufour, J-M. and M. Taamouti. (2005). "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," Econometrica, 73(4), 1351–1365.

Dufour, J-M. and O. Torrès. (1998). "Union-intersection and sample-split methods in econometrics with applications to SURE and MA models," In D. E. A. Giles and A. Ullah, editors, .Handbook of Applied Economic Statistics., pp. 465.505. Marcel Dekker, New York.

Elliott, G., T. J. Rothenberg, and J. H. Stock. (1996). "Efficient tests for an autoregressive unit root," Econometrica 64(4), 813.836.

Friedman, B. M. and D. I. Laibson. (1989). "Economic implications of extarordinary movements in stock prices (with comments and discussion)," Brookings Papers on Economic Activity 20, 137.189.

Gil-Pelaez, J. (1951). "Note on the inversion theorem," Biometrika 38, 481.482.

Gorman, T. (2004). Applied Adaptive Statistical Methods," Society for Industrial and Applied Mathematics.

Hotta, L. K. and R. S. Tsay. (1998). "Outliers in GARCH processes," unpublished manuscript Graduate School of Business University of Chicago.

Imhof, J. P. (1961). "Computing the distribution of quadratic forms in normal variables," Biometrika 48, 419.426.

Jansson, M. (2005). "Point optimal tests of the null hypothesis of cointegration," Journal of Econometrics 124, 187.201.

King, M. L. (1988). "Towards a theory of point optimal testing (with comments)," Econometric Reviews 6, 169.255.

Lehmann, E. L. and C. Stein. (1949). "On the theory of some non-parametric hypotheses," Annals of Mathematical Statistics 20, 28.45.

Lehmann, E. L. (1958). "Significance level and power," Annals of Mathematical Statistics 29, 1167. 1176.

Lehmann, E. L. (1959). "Testing Statistical Hypotheses," New York. John Wiley.

Lehmann, E. L. and J. P. Romano. (2005). "Testing Statistical Hypothesis," Springer Texts in Statistics. Springer-Verlag, New York., third ed.

Minkiw, N. G. and M. Shapiro. (1986). "Do we reject too often? small sample properties of tests of rational expectations models, " Economic Letters 20, 139.145

Pratt, J. and J. Gibbons. (1981). "Concepts of Nonparametric Theory," New York. Springer Verlag.

Rousseeuw, P. J. (1983). "Regression Techniques with High Breakdown Point," The Institute of Mathematical Statistics Bulletin, 12, 155.

Rousseeuw, P. J. and V. J. Yohai. (1984)."Robust Regression by Means of S-Estimators," in Robust and Nonlinear Time Series Analysis, ed. by W. H. Franke, and D. Martin, pp. 256–272. Springer-Verlag, New York.

Rousseeuw, P. J. and A. M. Leroy. (1987). "Robust Regression and Outlier Detection," Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Sanathanan, L. (1974). "Critical power function and decision making," Journal of the American Statistical Association 69, 398.402.

Schwert, G. (1990). "Stock volatility and the crash of 87," The Review of Financial Studies 3(1), 77.102.

White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, " Econometrica 48, 817.838.

Wright, J. H. (2000). "Alternative variance-ratio tests using ranks and signs, " Journal of Business and Economic Statistics 18(1), 1.9.

Yohai, V. J. and R. H. Zamar. (1988). "High Breakdown Point Estimates of Regression by Means of the Minimization of an Efficient Scale," Journal of the American Statistical Association, 83, 406–413.

Conclusion générale

Dans cette thèse, nous traitons des problèmes d'économétrie en macroéconomie et en finance. D'abord, nous développons des mesures de causalité à différent horizons avec des applications macroéconomiques et financières. Ensuite, nous dérivons des mesures de risque financier qui tiennent compte des effets stylisés qu'on observe sur les marchés financiers. Finalement, nous dérivons des tests optimaux pour tester les valeurs des paramètres dans les modèles de régression linéaires et non-linéaires.

Dans le premier essai nous développons des mesures de causalité à des horizons plus grands que un, lesquelles généralisent les mesures de causalité habituelles qui se limitent à l'horizon un. Ceci est motivé par le fait que, en présence d'un vecteur de variables auxiliaires Z, il est possible que la variable Y ne cause pas la variable X à l'horizon un, mais qu'elle cause celle-ci à un horizon plus grand que un [voir Dufour et Renault (1998)]. Dans ce cas, on parle d'une causalité indirecte transmise par la variable auxiliaire Z. Nous proposons des mesures paramétriques et non paramétriques pour les effets rétroactifs (feedback effets) et l'effet instantané à un horizon quelconque h. Les mesures paramétriques sont définies en termes de coefficients d'impulsion (impulse response coefficients) de la représentation VMA. Par analogie avec Geweke (1982), nous définissons une mesure de dépendance à l'horizon h qui se décompose en somme des mesures des effets rétroactifs de X vers Y, de Y vers X et de l'effet instantané à l'horizon h. Nous montrons également comment ces mesures de causalité peuvent être reliées aux mesures de prédictibilité développées par Diebold et Kilian (1998). Nous proposons une nouvelle approche pour évaluer ces mesures de causalité en simulant un grand échantillon à partir du processus d'intérêt. Des intervalles de confiance non paramétriques, basés sur la technique de bootstrap, sont également proposés. Finalement, nous présentons une application empirique où est analysée la causalité à différents horizons entre la monnaie, le taux d'intérêt, les prix et le produit intérieur bruit aux États-Unis. Les résultats montrent que: la monnaie cause le taux d'intérêt seulement à l'horizon un, l'effet du produit intérieur bruit sur le taux d'intérêt est significatif durant les quatre premiers mois, l'effet du taux d'intérêt sur les prix est significatif à l'horizon un et finalement le taux d'intérêt cause le produit intérieur bruit jusqu'à un horizon de 16 mois.

Dans le deuxième essai nous quantifions et analysons la relation entre la volatilité et les rendements dans les données à haut-fréquence. Dans le cadre d'un modèle vectoriel linéaire autorégressif de rendements et de la volatilité réalisée, nous quantifions l'effet de levier et l'effet de la volatilité sur les rendements (ou l'effet rétroactif de la volatilité) en se servant des mesures de causalité à court et à long terme proposées dans l'essai 1. En utilisant des observations à chaque 5 minute sur l'indice boursier S&P 500, nous mesurons une faible présence de l'effet de levier dynamique pour les quatre premières heures dans les données horaires et un important effet de levier dynamique pour les trois premiers jours dans les données journalières. L'effet de la volatilité sur les rendements s'avère négligeable et non significatif à tous les horizons. Nous utilisons également ces mesures de causalité pour quantifier et tester l'impact des bonnes et des mauvaises nouvelles sur la volatilité. D'abord, nous évaluons par simulation la capacité de ces mesures à détecter l'effet différentiel de bonnes et mauvaises nouvelles dans divers modèles paramétriques de volatilité. Ensuite, empiriquement, nous mesurons un important impact des mauvaises nouvelles, ceci à plusieurs horizons. Statistiquement, l'impact des mauvaises nouvelles est significatif durant les quatre premiers jours, tandis que l'impact de bonnes nouvelles reste négligeable à tous les horizons.

Dans le troisième essai, nous modélisons les rendements des actifs sous forme d'un processus à changements de régime markovien afin de capter les propriétés importantes des marchés financiers, telles que les queues épaisses et la persistance dans la distribution des rendements. De là, nous calculons la fonction de répartition du processus des rendements à plusieurs horizons afin d'approximer la Valeur-à-Risque (VaR) conditionnelle et obtenir une forme explicite de la mesure de risque «déficit prévu» d'un portefeuille linéaire à plusieurs horizons. Finalement, nous caractérisons la frontière efficiente moyenne-variance dynamique d'un portefeuille linéaire. En utilisant des observations journalières sur les indices boursiers S&P 500 et TSE 300, d'abord nous constatons que le risque conditionnel (variance ou VaR) des rendements d'un portefeuille optimal, quand est tracé comme fonction de l'horizon h, peut augmenter ou diminuer à des horizons intermédiaires et converge vers une constante- le risque inconditionnel- à des horizons suffisamment larges. Deuxièmement, les frontières efficientes à des horizons multiples des portefeuilles optimaux changent dans le temps. Finalement, à court terme et dans 73.56% de l'échantillon, le portefeuille optimal conditionnel a une meilleure performance que le portefeuille optimal inconditionnel.

Dans le quatrième essai, nous dérivons un simple test de signe point optimal dans le cadre des modèles de régression linéaires et non linéaires. Ce test est exact, robuste contre une forme inconnue d'hétéroscedasticité, ne requiert pas d'hypothèses sur la forme de la distribution et il peut être inversé pour obtenir des régions de confiance pour un vecteur de paramètres inconnus. Nous proposons une approche adaptative basée sur la technique de subdivision d'échantillon pour choisir une alternative telle que la courbe de puissance du test de signe point optimal soit plus proche de la courbe de l'enveloppe de puissance. Les simulations indiquent que quand on utilise à peu près 10% de l'échantillon pour estimer l'alternative et le reste, à savoir 90%, pour calculer la statistique du test, la courbe de puissance de notre test est typiquement proche de la courbe de l'enveloppe de puissance. Nous avons également fait une étude de Monte Carlo pour évaluer la performance du test de signe "quasi" point optimal en comparant sa taille ainsi que sa puissance avec celles de certains tests usuels, qui sont supposés être robustes contre hétéroscédasticité, et les résultats montrent la supériorité de notre test.