

Análise de Séries Temporais com Recurso a Técnicas de Bases de Dados

César Garcia¹, João Correia Lopes²

¹ Ministério da Finanças, Portugal
cfgarcia@mail.telepac.pt

² Faculdade de Engenharia da Universidade do Porto,
Rua Dr. Roberto Frias, 4200-465, Porto, Portugal
jlopes@fe.up.pt
<http://www.fe.up.pt/~jlopes/>

Resumo A evolução do mercado electrónico e a crescente utilização da Internet como meio de negócio, levou à disponibilização em vários *sites Web* da possibilidade de aquisição ou venda de acções em tempo real (*online*).

Neste trabalho descreve-se uma ferramenta inserida na área dos sistemas OLAP que permite a um eventual investidor baseado apenas no conhecimento histórico do valor das cotações das acções (séries temporais) efectuar, com alguma segurança, previsões sobre o comportamento futuro de determinada acção e assim encontrar oportunidades de negócio em tempo real.

Usam-se técnicas de bases de dados para a análise de séries temporais tendo sido obtidos bons resultados em simulações efectuadas com valores reais de cotações de acções da Bolsa de Valores de Lisboa (BVL).

1 Introdução

A análise de séries temporais é fundamental no apoio à tomada de decisões e definição de estratégias, a nível empresarial, financeiro ou outro, e cada vez mais organizações, independentemente do ramo de actividade em que estão inseridas, usam previsões na fase de tomada de decisões. As empresas possuem grande quantidade de informação armazenada ao longo das últimas décadas em bases de dados operacionais e, por vezes, em grandes armazéns de dados usados para apoio à decisão. Na tomada de decisões, a necessidade de obtenção de uma resposta imediata (*online*) leva a que se tentem encontrar soluções que optimizem o tempo de resposta às interrogações nas bases de dados onde reside a informação. A disponibilização de uma ferramenta de análise que permita dar uma resposta em tempo útil a uma possível operação de investimento no mercado de acções da Bolsa é essencial e necessária para este tipo de negócio, já que este é um ramo de actividade económica onde a previsão se pode mostrar essencial para o sucesso de uma operação de investimento.

Neste trabalho pretende-se, através da implementação de técnicas para acesso e manipulação de dados, proceder à análise das cotações das acções das empresas

cotadas na Bolsa e, assim, encontrar oportunidades de negócio. Para tal, foram analisadas as séries temporais das cotações das acções, através do estudo do comportamento diário do seu valor e do factor de correlação entre elas e foram estudadas as tendências das séries temporais, através do cálculo das médias móveis (neste caso, a média móvel de 5 e 21 dias).

O tipo de análise efectuada até ao momento não engloba todos os parâmetros que julgamos necessários e que permitiriam prever, com maior certeza, como vai ser o comportamento das acções de uma empresa qualquer. Tal não é possível, uma vez que:

- Uma análise aprofundada exigiria conhecimentos técnicos e científicos da área de Economia, para além das técnicas das Bases de Dados onde se insere este trabalho;
- Existem vários factores internos às próprias empresas (dados económicos das empresas, estratégias de gestão, etc.) aos quais não é possível ter acesso;
- Existem factores externos que influenciam, ou podem influenciar, o comportamento não só de determinado tipo de acções, mas até da(s) própria(s) Bolsa(s) de Valores — exemplos disso são as quedas generalizadas das bolsas muitas vezes influenciadas por factores totalmente alheios ao seu normal funcionamento.

O objectivo deste trabalho centra-se na obtenção de uma ferramenta computacional que permita a realização de investimentos na bolsa de valores a um investidor que não possua grandes conhecimentos sob o funcionamento do mercado de acções, ajudando-o a minimizar o risco inerente a este tipo de investimentos. Para apoiar o investidor na tomada de decisão sobre a compra e venda de acções foi desenvolvida uma aplicação em Java, chamada KBVL, que acede aos dados de uma base de dados e apresenta a interface com o utilizador que se apresenta na Figura 1. Nesta aplicação é feita a análise do factor de correlação entre séries de acções e a análise das médias móveis para estudo das tendências que se possam verificar, dando assim a oportunidade ao investidor de se decidir pela compra ou venda de acções, por forma a obter mais valias financeiras.

Seguidamente neste artigo serão descritas as técnicas de análise efectuadas, será detalhada a concepção e desenvolvimento da aplicação KBVL, serão apresentadas as interrogações à bases de dados e as regras de análise utilizadas. Para terminar são apresentadas as conclusões e enumeram-se possíveis trabalhos futuros.

2 Análise de Séries Temporais

Uma série temporal é uma sequência de dados, normalmente espaçados igualmente no tempo, ordenada cronologicamente (anual, mensal, semanal, diária, horária, etc.). O objectivo do estudo de uma série temporal é tentar descobrir padrões nos dados históricos e extrapolar esses padrões para o futuro.

A grande vantagem de utilizar as séries temporais para a análise e previsão, é a de não haver necessidade de descobrir ou conhecer os factores que afectam o seu

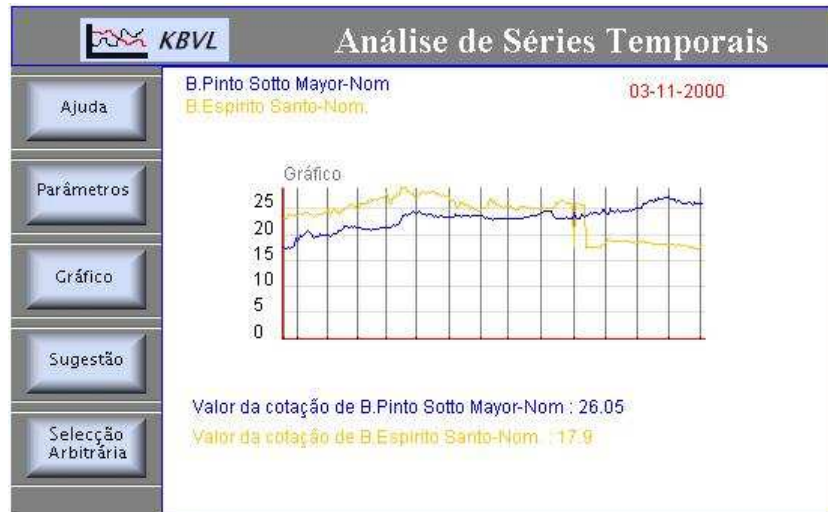


Figura 1. Interface da Aplicação KBVL

comportamento. Para este estudo, esta vantagem é ainda mais relevante se considerarmos que, nas variações a que estão sujeitas as cotações das acções, não é fácil identificar relações de causa-efeito tornando-se muito difícil, senão mesmo quase impossível, conhecer todos os parâmetros necessários a uma correcta análise. Por outro lado, o que se afigura importante é encontrar uma oportunidade de negócio independentemente da forma como esse objectivo é conseguido; ou seja, o importante é prever **o que acontece e não como acontece** e, sendo assim, o método da análise de séries temporais é adequado para fazer este tipo de previsões.

Dos vários componentes que constituem, ou podem constituir, a série, como por exemplo a sazonalidade, o ciclo, a aleatoriedade e a tendência, neste estudo apenas foi considerada a tendência, uma vez que o objectivo é apenas o de encontrar uma oportunidade de negócio em tempo real — comprar e vender — e não obter previsões para investimentos a médio ou longo prazo.

2.1 Factor de Correlação

O factor de correlação é talvez a mais poderosa ferramenta de análise que podemos utilizar no estudo das séries temporais. A correlação [CC89] mede a relação entre dois conjuntos de dados escalonados de forma a serem independentes da unidade de medida. O factor de correlação entre duas séries varia no intervalo compreendido entre os valores 1 e -1, podendo-se extrair do seu valor uma das seguintes conclusões:

- valor do coeficiente igual ou próximo de 1** — Os valores das duas séries estão correlacionados, isto é, sempre que se verifica o aumento do valor de uma série o valor da outra série também aumenta, e vice-versa.
- valor do coeficiente igual ou próximo de -1** — Os valores das duas séries estão inversamente correlacionados, isto é, sempre que se verifica o aumento do valor de uma série o valor da outra série diminui, e vice-versa.
- valor do coeficiente igual ou próximo de 0** — As séries não estão correlacionadas, ou seja, não podemos extrair qualquer conclusão, pois quando se verifica uma variação de valor numa série o valor da outra série varia aleatoriamente.

Neste trabalho o factor de correlação é usado para determinar pares de acções para as quais deve ser feita a análise de variação temporal, pois podem existir oportunidades de negócio. Assim, se por exemplo se verificar que existe uma correlação elevada³ entre as séries das acções da empresa A e da empresa B, fazendo ambas parte de uma dada carteira de acções (*portfolio*), e as acções da empresa A subiram de cotação e as acções da empresa B mantiveram a cotação, então a melhor atitude será comprar acções da empresa A e vender as acções da empresa B, a não ser que existam factores que eu conheça (por exemplo o lançamento de uma OPA (Oferta Pública de Aquisição) ou de uma OPT (Oferta Pública de Troca) sobre as acções da empresa A) que tenham levado a essa variação e que desaconselhem a operação.

Para escolha dos pares de acções a analisar podem ser considerados diversos critérios, como por exemplo:

- proceder ao estudo de todas as séries independentemente do ramo de actividade em que a empresa se encontra colocada, desde que se verifique um factor de correlação elevado;
- estudar apenas os pares de acções de empresas pertencentes ao mesmo ramo de actividade ou sector económico;
- considerar na análise apenas um de diversos períodos de tempo.

Atendendo à especificidade das séries analisadas neste trabalho, é de considerar que o passado recente é mais importante que o “histórico” [Sha98] e a variação ao longo de um período de tempo muito grande (anos) poderá servir apenas de mera indicação para uma análise mais aprofundada. Por outro lado, é aceite que as séries das empresas inseridas no mesmo ramo de actividade estão mais correlacionadas do que as séries de empresas pertencentes a ramos de actividades diferentes [CKL99], porque pertencendo ao mesmo ramo ou sector, estão sujeitas ao mesmo tipo de variação quando confrontadas com factores externos às próprias empresas. Dentro do mesmo ramo de actividade há que considerar

³ O significado exacto de séries altamente correlacionadas é muito subjectivo pois se tecnicamente um valor superior a 0,5 já significa que existe correlação entre as duas séries, no caso das séries de cotações de acções de empresas depende muito do potencial investidor, ou da sua capacidade de investimento, a quantificação deste valor. Na aplicação desenvolvida este valor é parametrizável.

também o tipo de actividade, isto é, se as empresas estão inseridas em grupos mais homogêneos e menos susceptíveis de variações bruscas (caso da banca ou seguros, em que se denota mais consistência e estabilidade nos dados) ou em grupos mais susceptíveis a variações bruscas (caso das novas tecnologias) e este é um facto a ter em consideração quando se pretende investir, tendo em vista a minimização do risco.

2.2 Médias Móveis

A média móvel de uma série de preços de acções é um indicador que nos mostra a soma do preço do valor das acções durante determinado período de tempo dividida pelo número de dias do período (a média aritmética nesse período), sendo recalculada para cada dia do período considerado [Geo00]. A denominação de móvel advém do facto de que a cada novo período de tempo, consideramos um novo dado e abandonamos o mais antigo. Como a quantidade de dados a analisar é sempre a mesma, leva a que se abandone sempre o primeiro dado da série — o mais antigo. Através da média móvel podemos constatar as mudanças de tendência e o desenvolvimento da série e assim, analisada cumulativamente com o factor de correlação, encontrar pontos de compra ou venda [Sha98,JS99,ZS02].

Existem diferentes tipos de médias móveis [Ach00]: simples (também conhecidas como aritméticas), exponenciais, triangulares, variáveis e ponderadas, podendo ser aplicadas a qualquer tipo de dados da série, ou seja, ao preço de fecho, ao preço máximo, ao preço mínimo, ao preço de abertura e ao volume de acções transaccionadas em determinado dia. A diferença entre a aplicação dos diferentes tipos de médias móveis prende-se, essencialmente, com o período de tempo que se pretende analisar. Enquanto, por exemplo, a média aritmética dá igual peso a cada elemento da série, a média exponencial ou ponderada dá maior peso aos elementos mais recentes da série numérica. Neste trabalho, para cálculo das médias móveis, consideram-se a média móvel de 5 e a média móvel de 21 dias tal como em [Sha98]. Este critério, que é discutível pois podem ser definidos outros períodos de tempo [Mur87], como por exemplo as médias móveis de 4, 9 e 18 dias, enquadra-se dentro dos parâmetros aceites de uma forma generalizada, representados na Tabela 1 [Ach00]:

Período	Média Móvel
Muito curto prazo	5-13 dias
Curto prazo	14-25 dias
Curto/Médio prazo	26-49 dias
Médio prazo	50-100 dias
Longo prazo	100-200 dias

Tabela 1. Períodos de tempo a considerar para cálculo da média móvel

Da análise das médias de 5 e 21 dias podemos deduzir que, quando se verifica que a primeira (a média de 5) cruza de baixo para cima sobre a segunda (a média de 21), tal significa que a tendência é de subida; neste caso foi detectado um ponto de compra e deve-se optar por comprar as acções objecto de análise. Inversamente, quando a primeira (a média de 5) cruza a segunda (a média de 21) de cima para baixo, existe tendência de descida e então devem-se vender as acções [Sha98,Mur87].

Face ao apuramento das diversas situações possíveis pode desenhar-se a árvore de decisão que vai posteriormente facilitar a implementação do algoritmo representado na Figura 2.

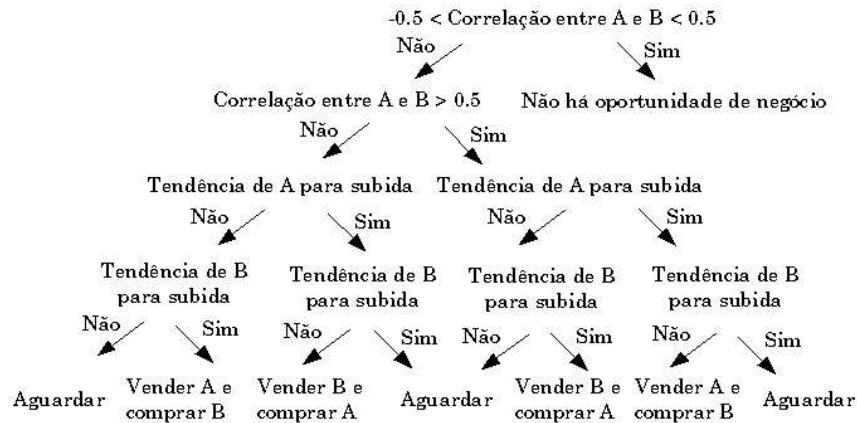


Figura 2. Árvore de decisão

Em conclusão, neste trabalho o factor de correlação permite-nos seleccionar pares de acções que apresentam uma variação entre elas mais ou menos similar e depois, através das médias móveis, podemos prever quando ocorrem oportunidades de negócio nestes pares.

3 Desenvolvimento da Aplicação KBVL

Para poder estudar o comportamento ou variação do valor das acções, é essencial e imprescindível possuir um histórico sobre o qual possamos trabalhar. Neste estudo, o histórico não é mais que o valor diário das cotações das acções na

Bolsa de Valores de Lisboa e, pontualmente, os valores dos dividendos⁴, bem como dos *splits*⁵.

A concepção da base de dados utilizada na aplicação desenvolvida teve em consideração os requisitos necessários à posterior manipulação dos dados (para cada empresa cotada na Bolsa de Valores) para o estudo das séries temporais, como sejam:

- Quantidade de acções transaccionadas em determinado período;
- Cotação máxima das acções em determinado período;
- Cotação mínima das acções em determinado período;
- Valor dos dividendos pagos e data do pagamento;
- Valor do factor de *Split* (caso exista);
- Data em que são anunciados os pagamentos de dividendos;
- Data em que são anunciados os *Splits*.

3.1 Armazenamento e Manipulação da Informação

Uma das características de um sistema de base de dados é permitir ao utilizador fazer interrogações por forma a obter a informação desejada, garantindo acesso eficiente a grandes quantidades de dados. Para tal é normalmente usada a linguagem SQL (*Structured Query Language*), sendo neste caso a resposta devolvida sob a forma de tabelas (conjuntos de tuplos). Por outro lado, a utilização de SQL impede, por definição de Modelo Relacional, que se tire vantagem da ordem das linhas, representando a ordem cronológica dos dados⁶. Por exemplo, não é possível, utilizando apenas SQL, responder a interrogações para:

- encontrar o lugar que ocupa determinado elemento na ordenação efectuada,
- realizar o cálculo de uma média móvel,
- realizar o cálculo do factor de correlação,
- proceder à comparação anual entre os valores das transacções.

Este tipo de insuficiências, associadas ao facto de que as bases de dados actuais cada vez possuem armazenadas maiores quantidades de informação, leva a que, apesar das suas vantagens, o modelo relacional se mostre bastante moroso quando o problema a tratar é relacionado com séries temporais. Nestes casos

⁴ Valor periodicamente pago (por acção) pela empresa aos seus accionistas proveniente dos lucros obtidos pela mesma.

⁵ Alteração do valor das acções após junção ou separação; por exemplo, a cada 5 acções passa a corresponder uma (factor de 0,2) levando o valor das acções a passar instantaneamente a 5x o valor actual; ou, a cada acção passam a corresponder 10 (factor de *split* 10) levando o valor das acções a passar instantaneamente a 1/10 do valor actual.

⁶ Pode-se evidentemente introduzir um atributo representando a ordem temporal dos dados e depois ordenar os dados a extrair (“`select <campo> from <tabela> order by <campo> ASC`”) mas esta operação é das mais demoradas. Podem também ser usados índices que melhorem o acesso aos dados mas a sua existência atrasa as inserções.

interessa não só aceder aos dados ordenados segundo critérios predeterminados mas, e essencialmente, poder posteriormente proceder à sua análise (cálculo do factor de correlação, médias móveis, etc.).

O KDB é um sistema de base de dados implementado em linguagem K⁷ e que reúne não só as características de um sistema tradicional de base de dados (com algumas particularidades como veremos mais à frente), com as características dos sistemas OLAP⁸ e OLTP⁹. A empresa que desenvolveu o sistema designou-o por sistema OLTAP.

Especificamente concebido para uma utilização sobre grandes volumes de dados, em que a informação precisa de ser rapidamente organizada, analisada, armazenada e distribuída (tal como em qualquer outro sistema de base de dados) KDB suporta como linguagem de acesso o ANSI-SQL II (vulgarmente denominado SQL 92), para além do KSQL. A linguagem KSQL [Kx 00c] foi definida especificamente para responder às exigências que são colocadas quer aos sistemas OLAP quer aos sistemas OLTP, podendo a informação ser originária de outros sistemas de bases de dados. A particularidade do KDB reside no facto de, contrariamente à maioria dos sistemas de bases de dados que se baseiam em tabelas, a estrutura dos dados ser o *arrable* (array-table). Existem outros sistemas OLAP com características semelhantes, como por exemplo o S-Plus¹⁰. O *arrable* é uma tabela em que a ordem dos dados pode ser facilmente explorada. Por exemplo, um campo de um registo da BD, pode conter uma série temporal, em que a ordem da série corresponde à ordem dos registos.

A exemplo do que acontece noutras bases de dados utilizando o SQL, com o KSQL também é possível efectuar as operações clássicas de agregação, junção, selecção, etc.. Para além destas operações “clássicas” é ainda possível definir funções para manipulação dos arrays, como por exemplo *reverse* (função que inverte a ordem dos elementos) ou *rank* (que nos dá o lugar ocupado por determinado elemento), etc. A importação de dados para KDB pode ser feita quer via ODBC, quer via JDBC [Kx 00b], isto no caso de se pretender analisar dados provenientes de outros sistemas, como por exemplo o MS Access que foi usado neste trabalho para armazenar a Base de Dados. Como qualquer outro sistema de base de dados, KDB também tem a sua própria estrutura organizativa da

⁷ K é uma linguagem de alto nível derivada (baseada) de APL, de NIAL e de LISP e que se caracteriza essencialmente pela forma como foram desenvolvidas e aperfeiçoadas as técnicas de manipulação de arrays.

⁸ OLAP (*online analytical processing*) são sistemas desenhados de forma a permitirem ao utilizador extrair e manipular de uma forma rápida grandes quantidades de dados armazenados em sistemas de bases de dados que contêm grande quantidade de informação.

⁹ OLTP (*online transaction processing*) são aplicações desenvolvidas tipicamente em versão cliente/servidor que visam facilitar a consulta e manipulação de grandes quantidades de dados de uma forma tão rápida quanto possível.

¹⁰ O S-Plus é um sistema derivado da linguagem S, desenvolvido na ATT Bell Laboratories e cujo desenvolvimento actualmente é da responsabilidade da MathSoft Inc.. Informação detalhada pode ser consultada em: <http://mathsoft.com/splus/> e <http://math.umbc.edu/~arghya/Splus/splus.html>

informação, sendo as tabelas armazenadas sob a forma CSV (*comma separated value*) que é um formato de troca típico.

O KSQL é muito parecido com o SQL, quer no tipo de operações que permite realizar e que atrás já foram referidas, quer na construção de instruções. A principal diferença consiste em que, em KDB as tabelas são ordenadas, ou seja, ao instanciar uma base de dados em KDB (`load <base_de_dados>`) as tabelas são automaticamente ordenadas e o KSQL explora esse facto. Socorrendo-se de funções já implementas na própria linguagem ou permitindo a sua implementação em linguagem K e em ficheiros com a extensão “.t”, perguntas à base de dados que em SQL não são possíveis, em KSQL fazem-se de uma forma directa, como veremos de seguida.

4 Interrogações à base de dados

A análise e estudo das séries temporais é, neste trabalho, baseada essencialmente no estudo do comportamento das séries através do respectivo coeficiente de correlação e das médias móveis.

Para calcular o coeficiente de correlação é necessário seguir os seguintes passos:

1. Saber quantos registos diários existem na base de dados respeitantes a cada uma das empresas (acções) a analisar, já que o factor de correlação apenas pode ser calculado entre séries com o mesmo número de valores;
2. Realizar a pergunta à base de dados para obter a série respeitante à primeira empresa¹¹;
3. Repetir o procedimento anterior para a segunda empresa a analisar;
4. Para cada uma das séries calcular o respectivo desvio padrão;
5. Calcular a covariância das séries;
6. Calcular o factor de correlação.

Como os cálculos estatísticos referidos nos pontos 4, 5 e 6 não fazem parte da norma SQL-92, não é possível realizá-los através de instruções em SQL. Assim, para efectuar o cálculo do desvio padrão e da covariância, é necessário implementar funções específicas com recurso a outras linguagens que devolvam esse valor, recebendo como argumento um vector que contenha toda a série a analisar.

Utilizando KSQL, os seis passos descritos no ponto anterior reduzem-se a apenas quatro evitando-se ainda o ajustamento dos vectores por forma a que contenham o mesmo número de valores:

1. Saber quantos registos diários existem na base de dados respeitantes a cada uma das empresas (acções) a analisar;

¹¹ Apesar de em SQL se poder realizar a pergunta apenas uma vez e retornar as duas séries, o que aumentaria sem qualquer dúvida a rapidez de execução já que evitava uma nova consulta à base de dados, o facto de nem sempre o número de registos relativos a cada série ser igual implica um tratamento diferenciado para cada uma das séries.

2. Realizar a pergunta à base de dados para obter a série respeitante à primeira empresa;
3. Repetir o procedimento anterior para a segunda empresa a analisar;
4. Calcular o factor de correlação.

Neste ponto a vantagem de utilização de KSQL em vez de SQL é notória: enquanto em SQL temos necessidade de perguntar à base de dados quais os valores das séries e posteriormente, através da implementação de um ciclo, ler todo o conjunto de tuplos da resposta (*recordset*) e carregar um vector de forma a que o número de elementos dos dois vectores (que contêm as séries) seja igual, em KSQL todo este processo se faz directamente.

De seguida são apresentados fragmentos do programa em Java que exemplifica este processo.

Para saber qual o mínimo de registos na base de dados comuns às séries da empresa A e da empresa B, pode-se perguntar directamente através da instrução:

```
String minimo= "select min x, Descricao from (
    select count by Descricao from Cotacoes
    where Descricao in ('" + empA + "',''" + empB + "''))";
Object[]rmin = (Object[])c.k(minimo); // executa a pergunta
Minimo = (((int[])dmin[0])[0]); // valor mínimo de registos
```

Como é óbvio é preciso saber quantos registos existem no histórico relativamente a cada empresa. Essa pergunta é muito semelhante à que seria formulada em SQL:

```
String all= "select count by Descricao from Cotacoes
    where Descricao in ('" + empA + "',''" + empB + "''))";
Object[]cor11 = (Object[])c.k(all); // executa a instrução
Object[]d11 = (Object[])cor11[1]; // valor total de registos
drops = (((int[])d11[1])[1] - Minimo); // número de registos a ignorar
```

Na posse destes dois valores (o menor e o valor total dos registos), é possível interrogar a base de dados de forma a serem devolvidos apenas os registos necessários para a análise. Como o objectivo da pergunta é saber os valores a partir dos quais vão ser calculadas as médias móveis de 5 e 21 dias, neste passo é possível explorar as características da linguagem e proceder simultaneamente à execução da pergunta e ao cálculo daquelas duas médias:

```
String query= "select by " +drops+ "drop Datad, " +drops+ " drop Preco, "
    + drops + " drop 5 avgs Preco, " + drops + " drop 21 avgs Preco
from Basc where Descricao = '" + empB + "'";
```

Nesta instrução a palavra reservada *drop* significa que o servidor ignora os primeiros *drops* registos e que as médias devem ser calculadas apenas sobre os registos devolvidos, ou seja, não é acumulada na soma, para efeitos de cálculo da média, a soma dos registos que foram ignorados.

Para o cálculo do factor de correlação temos necessidade de nos socorrer de funções previamente implementadas em linguagem K [Kx 00a] e, da mesma forma que é executada uma instrução (em SQL ou KSQL), pode-se chamar directamente a função pretendida:

```
Object xy= correlac.k("cor", (double[])d2[1], (double[])d1[1]) em que:
```

- `cor` é o nome da função previamente definida em linguagem K para cálculo do factor de correlação; e
- `(double[])d2[1], double[])d1[1]` são os dois vectores que contêm as séries temporais, sendo depois o valor dado directamente por:

```
correlacao+=(Double.valueOf((xy).toString()).doubleValue());
```

5 Exemplo de utilização da aplicação KBVL

Nesta aplicação considerou-se que apenas seriam analisados os pares de acções em que o factor de correlação entre as séries fosse maior que 0,5 ou menor que -0,5. Sempre que se verifique esta situação, é então analisada a tendência de cada série (com recurso às médias móveis de 5 e 21 dias) e, caso se verifique uma oportunidade de negócio é dada a sugestão de compra ou venda. Como se pode ver na Figura 3, o utilizador pode escolher um factor de correlação mínimo para os pares de cotações que pretende ver tratados.

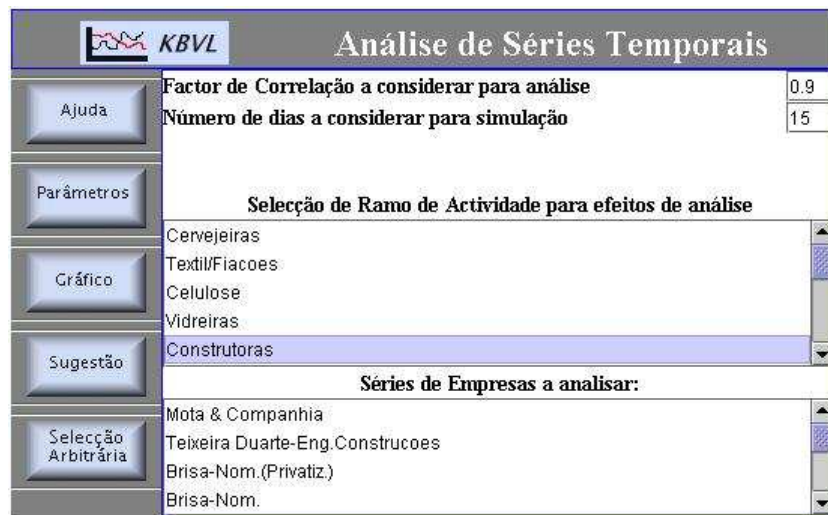


Figura 3. Utilização da Aplicação KBVL

Dos exemplos de simulação de investimento efectuados, em grande número, durante vários meses, podemos concluir que:

- Para **séries de empresas inseridas dentro do mesmo ramo de actividade altamente correlacionadas positivamente** foram obtidos resultados finais francamente bons em termos de oportunidades de negócio;

- Para **séries de empresas inseridas dentro do mesmo ramo de actividade altamente correlacionadas negativamente** foram igualmente obtidos resultados finais francamente bons em termos de oportunidades de negócio;
- Para **séries altamente correlacionadas positivamente mas de empresas de ramos de actividade distintos** os resultados finais indicam que, frequentemente, se chegam a maus negócios;

A análise dos resultados das simulações efectuadas indica claramente que a probabilidade de sucesso é de facto maior quando as empresas a analisar estão inseridas dentro do mesmo ramo de actividade. Contudo, em certos ramos de actividade, o risco é muito maior dada a variação desconforme que se verifica nas séries. Estas variações poderão ter várias explicações que fogem da simples análise através do estudo do factor de correlação e tendência, tornando-se necessário ter um conhecimento exterior sobre os motivos que levaram a estas variações.

6 Conclusão e Trabalho Futuro

Em que acções devo investir? Esta é a pergunta que possivelmente todos gostariam de ver respondida, pois estaríamos todos ricos!

Como anteriormente foi referido, todo o trabalho foi desenvolvido tendo em vista a comparação entre duas séries de valores de acções diferentes, em que as empresas a analisar estão inseridas dentro do mesmo ramo de actividade e, com base nessa comparação tentar encontrar oportunidades de negócio. Este é pois um campo por onde poderá evoluir este trabalho, sendo que os objectivos terão que ser necessariamente diferentes.

Uma outra hipótese de desenvolvimento futuro, e que não foi implementada neste trabalho, prende-se com a forma como foram estudadas as séries das acções, isto é, neste trabalho pegou-se em toda a série e aplicaram-se directamente os métodos de análise — factor de correlação e médias móveis (e através destas as tendências) — sem ter a preocupação de analisar os valores constantes da série. Num estudo mais aprofundado poder-se-iam aplicar conjuntamente outras técnicas de análise das séries, o que poderia conduzir à produção de soluções um pouco diferentes.

Referências

- [Ach00] Steven B. Achelis. *Technical Analysis from A to Z*. 2000.
- [CC89] S.C. Chapra and R.P. Canale. *Numerical Methods of Engineers*. McGraw-Hill, New York, 1989.
- [CKL99] Louis K.C. Chan, Jason Karseski, and Josef Lakonishok. On portfolio optimization: Forecasting covariances and choosing the risk model. *The review of Financial Studies*, 12(5), 1999.
- [Geo00] Geocities. Fundamentos de análise técnica e seus instrumentos. URL: <http://www.geocities.com/WallStreet/2500/przpanap.htm>, May 2000.

- [JS99] Kaippallimalil J. Jacob and Dennis Shasha. Fintime - a financial time series benchmark. *ACM SIGMOD Record*, 28(1):42–48, March 1999.
- [Kx 00a] Kx Systems. *The C - K Interface*, 2000.
- [Kx 00b] Kx Systems. *KDP Programming Guide*, 2000.
- [Kx 00c] Kx Systems. *KSQL Reference Manual*, 2000.
- [Mur87] John J. Murphy. *Technical Analysis of the Futures Markets*. New York Institute of Finance-70, Pine Street-New York, New York 10270, 1987.
- [Sha98] Dennis Shasha. Time series in finance: the array database approach. Courant Institute of Mathematical Sciences, Department of Computer Science, New York University, 1998.
- [ZS02] Yunyue Zhu and Dennis Shasha. Statstream: Statistical monitoring of thousands of data streams in real time. In *VLDB'2002, Very Large Databases, Hong Kong, China*, pages 358–369, 2002.